Beiträge zur Schätzung des jährlichen durchschnittlichen Schadenbetrages in der Motorfahrzeughaftpflichtversicherung

INAUGURAL-DISSERTATION

der

Philosophisch-naturwissenschaftlichen

Fakultät der Universität Bern

zur

Erlangung der Doktorwürde

vorgelegt von

ROLF ECKERT

von Bern

VERLAG P. G. KELLER - WINTERTHUR 1957

Beiträge zur Schätzung des jährlichen durchschnittlichen Schadenbetrages in der Motorfahrzeughaftpflichtversicherung

INAUGURAL-DISSERTATION

der

Philosophisch-naturwissenschaftlichen

Fakultät der Universität Bern

zur

Erlangung der Doktorwürde

vorgelegt von

ROLF ECKERT

von Bern

VERLAG P. G. KELLER - WINTERTHUR 1957

Von der philosophisch-naturwissenschaftlichen Fakultät auf Antrag des Herrn Prof. Dr. A. Alder angenommen.

Bern, den 5. März 1955

Der Dekan:

W. Scherrer

MEINEM VATER GEWIDMET

Es ist mir ein Bedürfnis, an dieser Stelle meinen verehrten Lehrern, insbesondere den Herren

Prof. Dr. A. Alder
Prof. Dr. W. Scherrer
Prof. Dr. H. Hadwiger
Prof. Dr. H. Greinacher
Prof. Dr. A. Mercier
Prof. Dr. W. Nef
Prof. Dr. W. Michel
Prof. Dr. W. Wegmüller
Prof. Dr. W. Koenig

für ihren vorzüglichen Unterricht und das mir während meines Studiums erwiesene Wohlwollen aufrichtig zu danken.

Ganz besonders bin ich Herrn Prof. Dr. A. Alder, unter dessen umsichtiger Leitung die vorliegende Arbeit entstanden ist, sowie Herrn Dr. F. Bichsel von der Berner Allgemeinen Versicherungs-Gesellschaft für seine wertvollen Ratschläge und praktischen Hinweise zu bleibendem Dank verpflichtet.

INHALTSVERZEICHNIS

0 *Einleitung*

01 Vorbemerkungen 1
02 Allgemeine Problemstellung 3

1 *Analytische Lösungsversuche*

10 Allgemeines 7
11 Die bisherigen analytischen Ansätze für die Verteilung der Schäden in der Autohaftpflichtversicherung 8
12 Die lognormale Verteilung 12

121 Hypothese 12
122 Allgemeines über die Lognormalverteilung; Literaturhinweise 12
123 Die wichtigsten theoretischen Kenngrössen der Lognormalverteilung; Zusammenhänge mit der Gauss'schen Normalverteilung 13
124 Bestimmung der Parameter ξ und σ 16
125 Schätzwerte für den Durchschnitt $\mu(y)$ und die Streuung $\sigma^2(y)$ 16
126 Anwendung. Überprüfung der Hypothese von 121 anhand unseres Beobachtungsmaterials 18

13 Die PEARSON Typ III-Verteilung 22

131 Allgemeines 22
132 Hypothese 22
133 Theoretische Kenngrössen 23
134 Bestimmung der Parameter α und γ 24
135 Durchschnitt $\mu(y)$ der Originalgrundgesamtheit; seine Schätzung durch die transformierten Grössen $x=\log y$ 26

136 Überprüfung der Hypothese 132 am Untersuchungsmaterial 27

14 Folgerungen aus den analytischen Lösungsversuchen . . . 29

2 *Grundzüge einer neuen Methode zur Bestimmung des durchschnittlichen Schadenbetrages*

21 Die grundlegende Idee. Das geometrische Mittel als charakteristische Grösse 31

22 Homogenitätsteste zur Bestätigung der Hypothese der Zeitinvarianz des Mittelwertquotienten Q^1 33

221 Theoretische Grundlage 33

222 Praktische Auswertung 34

23 Die Häufigkeitsverteilung des geometrischen Mittels $y_G{}^1$ bezw. $\log y_G{}^1$ 37

24 Bestimmung des Mittelwertquotienten Q und dessen Streuung mittels Stichprobenbildungen. 39

25 Die Schätzung des durchschnittlichen Schadenbetrages M^1 und dessen Streuung. 42

3 *Praktische Anwendungen*

31 Die numerische Ermittlung des Mittelwertquotienten Q und dessen Streuung. 44

311 Bildung der Stichproben aus dem gesamten Beobachtungsmaterial: Autohaftpflichtschäden „BERNER ALLGEMEINE", 1938, 1947–1951 . . 44

312 Normalitätsteste für die in 24 getroffenen Hypothesen 45

313 Gegenüberstellung der nach den verschiedenen Ansätzen in 24 berechneten Mittelwertquotienten . . . 46

314 Zusätzliche Bemerkung 48

32 Die numerische Bestimmung des durchschnittlichen Schadenbetrages M^1 und dessen Streuung $\sigma^2(M^1)$. 48

33 Bestimmung eines Schadensatzes als weitere Anwendung 50

34 Zusammenfassung und Ausblick 51

4 *Anhangtabellen* 53

5 *Literaturverzeichnis* 56

0 Einleitung

01 VORBEMERKUNGEN

Es ist schon von verschiedenen Seiten auf die Tatsache hingewiesen worden, dass mathematische Methoden in fast allen Versicherungszweigen, die nicht in das Gebiet der Lebensversicherung fallen, nur langsam entwickelt werden und zur Anwendung gelangen. RIEBESELL [1] * spricht dann von einer mathematischen Methode in der Sachversicherung, wenn bei Untersuchungen in diesen Versicherungsbranchen Methoden der Wahrscheinlichkeitsrechnung oder der mathematischen Statistik beigezogen werden. JECKLIN [2] bemerkt dazu mit Recht, dass dadurch die Sachversicherungsmathematik begrifflich viel enger definiert sei als die Lebensversicherungsmathematik; denn hier werden die Ergebnisse der Wahrscheinlichkeitstheorie nur gerade bei der Herleitung der Grundlagen verwendet. Der eigentlichen Lebensversicherungsmathematik wie Prämien- und Reservenberechnung liegen eher bilanztechnische als mathematische Verfahren zugrunde.

Vergleichen wir nur schon die Zahl der veröffentlichten Arbeiten mathematischen Charakters über Lebensversicherungsfragen einerseits und von Sach- oder Vermögensversicherungsproblemen anderseits, so können wir wohl von einem Missverhältnis sprechen. Trotz des starken Aufschwunges der mathematischen Statistik, und trotzdem uns gerade in jüngster Zeit durch die kollektive Risikotheorie ein äusserst nützliches Werkzeug in die Hand gegeben wird, hat sich eigentlich, besonders bezüglich der praktischen Auswirkungen, nicht viel an diesem Zustand geändert.

Es besteht auch nicht die Absicht, hier all die vorgebrachten Gründe für dieses Versagen aufzuführen; darüber ist genügend diskutiert und geschrieben worden, dazu sei auf die zitierte Arbeit [2] und auf die Einleitung der Dissertation von LEEPIN [3] verwiesen, der

* Die in eckige Klammern gesetzten Zahlen verweisen auf das Literaturverzeichnis.

darin eine gute Übersicht über diesen von verschiedenen Fachleuten geführten Meinungsstreit wiedergibt. Überdies finden wir hier auch Literaturangaben der grundlegendsten Arbeiten über die Sachversicherungsmathematik.

Eines ist klar: solange sich die bisherigen Verfahren zur Beurteilung der verschiedenen Risiken und zur Berechnung von Tarifen in der Schadenversicherung bewähren, solange dürfte das „menschliche Beharrungsvermögen" gross sein; diese vorwiegend intuitiv und empirisch begründeten Methoden haben sicherlich auch ihre guten und praktischen Seiten. Doch sobald sich in einer Risikogattung eine ungünstige Entwicklung abzeichnet, steigt das Bedürfnis nach genaueren und zuverlässigeren Schätzungen. Hier liegt denn auch ein fruchtbares Arbeitsgebiet für den Theoretiker zur Entwicklung mathematisch-statistischer Methoden vor.

In dieser Beziehung bietet gerade die *Motorfahrzeug-Haftpflichtversicherung* ein dankbares Feld für derartige Untersuchungen. Diese Branche hat sich in den letzten Jahren durch die sprunghafte Zunahme des Motorverkehrs überall gewaltig entwickelt. Während für die meisten anderen Schadenversicherungszweige der Grundsatz einer gewissen Stabilität in der Risikenzusammensetzung im Laufe der Jahre erfüllt ist, kann man dies vom Autohaftpflichtgeschäft nicht unbedingt behaupten. Dies geht schon aus der zweimaligen Tarifrevision innerhalb der kurzen Zeitspanne 1947–1951 in unserem Land hervor. Anderseits erlaubt schon das Erfahrungsmaterial einer einzigen Gesellschaft die Anwendung neuer stichprobentheoretischer Verfahren, da bei einer bestimmten Grösse des Portefeuilles die Erfüllung des Gesetzes der grossen Zahlen gewährleistet sein dürfte.

Es ist im Rahmen dieser Arbeit nicht möglich, eine umfassende mathematische Darstellung der Motorfahrzeughaftpflichtversicherung zu entwickeln. Die Betrachtungen werden auf die Analyse einer wichtigen, charakteristischen Grösse, des *durchschnittlichen oder mittleren Schadenbetrages*, beschränkt. Diese Grösse bestimmt zusammen mit der Schadenhäufigkeit die Prämie jeder Sach- und Vermögensversicherungsart. In der Haftpflichtversicherung im besonderen ist der durchschnittliche Schadenbetrag eine von Jahr zu Jahr mit grossen Zufallsschwankungen behaftete Grösse, da wir es hier nicht mit einer Summenversicherung zu tun haben; die Leistungspflicht des Versicherers

im Schadenfalle ist lediglich durch eine sogenannte Höchst-Garantiesumme begrenzt. Eine nähere Untersuchung lohnt sich hier sowohl vom theoretischen als auch vom praktischen Standpunkt aus.

Die Beschränkung unseres Problemkreises bedingt hauptsächlich eine eingehende Analyse der Schäden und ihrer Häufigkeiten innerhalb passend gewählter Intervallbreiten. Zur Bestätigung oder Verwerfung aufgestellter Hypothesen und theoretischer Entwicklungen stand Erfahrungsmaterial aus Autohaftpflichtschäden der „Berner Allgemeinen" uneingeschränkt zur Verfügung; der Gesellschaft sei an dieser Stelle für die grosszügige Hilfe bestens gedankt.

Es ist das Ziel der vorliegenden Studie aufzuzeigen, wie auch in Versicherungsgebieten ausserhalb der Lebensversicherung neue mathematisch-statistische Verfahren entwickelt oder schon bestehende Methoden sinngemäss übertragen werden können, deren Auswertung sich auch in der Praxis lohnt. Es wurde oben verschiedentlich angetönt, dass in diesen Zweigen der Versicherung noch viel fruchtbarer, aber unbeackerter Boden vorliegt. In diesem Sinne möchten die folgenden Ausführungen einen Beitrag bilden.

02 ALLGEMEINE PROBLEMSTELLUNG

Es liege das *vollständige Geschäftsergebnis eines bestimmten Jahres i* vor, d.h. aus dem zu untersuchenden Autohaftpflicht-Portefeuille kennen wir:

n^i: Zahl der Schäden im Jahre i

$y_1^i, y_2^i, \ldots, y_{n^i}^i$: einzelne Schäden im Jahre i, geordnet nach der Höhe ihres Schadenbetrages.

Unsere zwei Problemstellungen lauten dann wie folgt:

1. *Wie kann aus diesem zahlenmässig beschränkten und grossen Zufallsschwankungen unterworfenen Material auf einen möglichst zuverlässigen durchschnittlichen Schadenbetrag M^i des Jahres i geschlossen werden, der überdies geschäftsspezifischen Charakter haben soll?*
2. *Läßt sich eine Wahrscheinlichkeitsaussage über die Genauigkeit des so geschätzten Durchschnittes M^i machen.*

Die Art der Schadenregulierung und die Grundsätze bei der Auswahl der Risiken sind von Gesellschaft zu Gesellschaft verschieden;

sie beeinflussen die Grösse M^i entscheidend. Deshalb wünschen wir einen „geschäftsspezifischen" Durchschnittswert.

Die einfachste und bei den meisten übrigen Schadenversicherungszweigen übliche Methode, einen Schätzwert für M^i und für dessen Streuung zu finden, besteht darin, aus der Gesamtschadensumme

$$S^i = \sum_{j=1}^{n^i} y_j{}^i$$

den Durchschnitt $\bar{y}^i$ zu bilden:

$$E(M^i) = \bar{y}^i = \frac{S^i}{n^i} \tag{1}$$

Nach dem zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie wird dann angenommen, dass die zufällige Variable $\bar{y}^i$ *angenähert normalverteilt liegt* mit einem Durchschnitt nach (1) und einer mittleren quadratischen Abweichung von

$$\sigma^i(M^i) = \frac{\sigma^i(y)}{\sqrt{n^i}} \text{ wobei } \sigma^i(y) = \sqrt{\frac{1}{n^i - 1} \sum_{j=1}^{n^i} (y_j{}^i - \bar{y}^i)^2} \tag{2}$$

wodurch wir den Fehler des geschätzten M^i nach dem üblichen Verfahren beurteilen können.

Praktisch müssten wir passend gewählte, möglichst äquidistante Klassenintervalle bilden, da die Zahl der Schäden n^i in der Regel ziemlich gross ist; ohne diese Massnahme wäre vor allem die Streuungsberechnung nach (2) zu umständlich. Hier aber begegnen wir schon einer *ersten Schwierigkeit*. Lässt sich eine solche Gruppenbildung nach der Grösse der Schäden bei Summenversicherungen leicht durchführen, so ist dies bei Haftpflichtversicherungen ziemlich problematisch wegen der grossen Schwankungsbreite der Schadenbeträge. In unserem Autohaftpflichtportefeuille liegen diese praktisch zwischen 10 und 100000 Franken und eventuell mehr, da wir *Sach- und Körperschäden gesamthaft* untersuchen. Eine Intervallbreite von 1000 Franken z.B. wäre für die kleinen Schäden zu gross und für die grossen zu klein, um die der empirischen Verteilung zugrunde liegenden Gesetzmässigkeiten überhaupt noch aufdecken zu können. Im nächsten Kapitel werden wir sehen, wie wir dieses Hindernis mittels einer logarithmischen Klasseneinteilung zu überbrücken vermögen, wodurch natürlich die analytische Betrachtungsweise und die vorzu-

nehmenden Grenzübergänge gegenüber den üblichen statistischen Schlüssen geändert werden müssen.

Wir stellen schon hier fest, dass ein nach (1) geschätztes M^1 (für das Folgende als „direkter" Durchschnitt $\bar{y}^1$ bezeichnet) eine praktisch zu grosse Streuung (2) aufweist, um noch eine gültige, statistisch gesicherte Aussage anhand eines Jahresergebnisses zuzulassen.

Die *Hauptschwierigkeit* liegt aber bei den innerhalb eines Jahres auftretenden *grossen Schäden*, die ersichtlich den Durchschnittsschaden M^1 bei der grossen Variationsbreite entscheidend beeinflussen. Im Vergleich zu den häufigen und gesetzmässigen kleinen Schäden sind diese grossen Schäden selten und hauptsächlich sehr zufallsbedingt. Das *Kernproblem* lässt sich daher kurz wie folgt ausdrücken:

Wie kann aus dem gesetzmässigen Verlauf der Bagatellschäden auf die Häufigkeit des Auftretens der grossen Schäden, die fast ausschliesslich die Höhe des durchschnittlichen Schadenbetrages bestimmen, geschlossen werden?

Unwillkürlich denkt man nun daran, einen grösseren Zeitraum zusammenzufassen, um möglichst auch bei den grossen Schäden genügendes Beobachtungsmaterial zu gewinnen. Dazu müssten wir eine Periode von 5 Jahren und mehr berücksichtigen; die Homogenität des Untersuchungsmaterials wäre aber dadurch wegen der Kurzfristigkeit der Autohaftpflicht-Verträge und wegen der allgemeinen Teuerungserscheinungen sehr in Frage gestellt. Anderseits führte uns dies eher von unserer Problemstellung weg, da wir ja gerade aus den Zahlen *eines Jahres* einen möglichst wirklichkeitsgetreuen durchschnittlichen Schadenbetrag M^1 bestimmen wollen, um aus diesem u.a. auch die allgemeine Entwicklungstendenz von Jahr zu Jahr beurteilen zu können, was für eine vernünftige Prämien- und Rückversicherungspolitik unerlässlich ist.

Das Naheliegendste ist, eine *analytische Häufigkeitsfunktion der Schadenbeträge*, die die empirische Verteilung eines Jahres möglichst gut annähert, zu suchen. Daraus kann ein theoretischer Durchschnitt und dessen Streuung durch Integralbildung bestimmt und mit Hilfe von Stichprobenmethoden deren „beste" Werte geschätzt werden. Im 1. Kapitel werden die in diesem Sinne schon gemachten Ausgleichsvorschläge auf ihre Verwendbarkeit hin überprüft (11) und zwei weitere analytische Lösungsversuche studiert (12 und 13). Alle diese mathematischen Formulierungen der Verteilung der Schäden

nach ihren Beträgen werden für unsere Zwecke nicht befriedigen, sie geben den Zufallscharakter der grossen Schäden und damit denjenigen von M^1 zu wenig gut wieder (14).

Es wird deshalb im *2. Kapitel* ein *neues*, *direkteres Verfahren* entwickelt. Dessen Bewährung wird anhand des Beobachtungsmaterials der Jahre 1947–1951 gezeigt, und weitere Anwendungsmöglichkeiten der dargelegten Theorie werden angetönt (3. Kapitel).

1 Analytische Lösungsversuche

10 ALLGEMEINES

Es wird grundsätzlich die *Existenz einer hypothetischen Grundgesamtheit* gefordert, deren mathematische Form die analytische Häufigkeitsverteilung der Schadengrössen y liefert; *y ist demnach eine zufällige Variable*. Die der empirischen Verteilung der Schäden angepasste kontinuierliche *Häufigkeitsfunktion* $g(y)$ definiert diese Grundgesamtheit; $g(y)dy$ bezeichnet dann die Wahrscheinlichkeit, dass ein beliebiger Schaden einen Betrag zwischen y und $y + dy$ aufweist. Das Integral

$$G(y) = \int_0^y g(y)dy$$

bezeichnet weiter die Wahrscheinlichkeit, dass dieser Schaden den Betrag y nicht übersteigt. $G(y)$ ist die *Verteilungsfunktion* der zufälligen Variablen y.

Die eingetretenen Schäden können dann als Zufallsstichproben aus dieser hypothetischen Grundgesamtheit $g(y)$ angesehen werden. Wahrscheinlichkeitstheoretisch gesprochen haben wir ein Rückschlussproblem: wie kann aus den beobachteten Stichprobendaten auf die unbekannte Grundgesamtheit geschlossen werden? Angenommen, wir hätten deren mathematisches Gesetz $g(y)$ gefunden, stellt sich dann die für unser Problem besonders wichtige Frage nach der besten Bestimmung der Parameter und der Kenngrössen wie Durchschnitt und Streuung dieser Häufigkeitsfunktion. Diese Fragestellungen gehören in das weitläufige Gebiet der Stichproben- und Schätzungstheorie, die in den letzten Jahrzehnten stark entwickelt worden ist.

Das Problem einer geeigneten mathematischen Ausgleichsfunktion für die Autohaftpflichtschäden wurde von einigen Forschern untersucht, ohne dass dabei auf eine tiefere, wahrscheinlichkeitstheoretische Betrachtungsweise im obigen Sinn eingegangen wurde; man suchte

einfach nach einer Häufigkeitsfunktion, die die beobachteten Schäden möglichst gut annähert, und zwar in der gesamten Intensitätsbreite von den kleinsten bis zu den grössten Schäden. Wir werden im Folgenden diese schon gemachten Studien erwähnen und auf ihre Verwendbarkeit für unsere Problemstellung hin überprüfen (11). In den zwei weiteren Abschnitten (12 und 13) werden Lösungsansätze etwas eingehender studiert, denen eine logarithmische Klasseneinteilung zugrunde gelegt ist, während die in 11 zusammengestellten Häufigkeitsfunktionen eine willkürlich gewählte, äquidistante Gruppierung der Schäden voraussetzen (siehe dazu auch Bemerkung in 02).

11 DIE BISHERIGEN ANALYTISCHEN ANSÄTZE FÜR DIE VERTEILUNG DER SCHÄDEN IN DER AUTOHAFTPFLICHTVERSICHERUNG

a) Myslivec [4], aus dessen Arbeit auch die folgende Zusammenstellung der bisher erschienenen Beiträge teilweise übernommen wurde, setzte selbst die bekannte, einfache Pareto-Funktion an:

$$g(y) = a y^{-b} \tag{1}$$

Er begründet die Anwendbarkeit dieser zur Darstellung von Häufigkeiten in der Wirtschaft (z.B. Einkommensverteilungen) vielfach gebrauchten Funktion hauptsächlich damit, dass die Institution des Versicherungswesens von den gleichen ökonometrischen Gesetzmässigkeiten beherrscht sei wie die ganze Volkswirtschaft; alle Erscheinungen in den wirtschaftlichen Verhältnissen der Bevölkerung und der wirtschaftlichen Betriebe spiegle sich auch in allen Versicherungszweigen wider. Er verweist auf die erfolgreiche Anwendung der Pareto-Funktion für die Risikosummenverteilung in der Lebens- und Brandschadenversicherung (für diesbezügliche Literaturangaben siehe Originalarbeit [4]). Es sei nur natürlich, diese Funktion auch in der Autohaftpflichtversicherung, die doch ein Bestandteil der allgemeinen Volkswirtschaft bilde, anzusetzen.

Auf Grund einer Untersuchung am Zahlenmaterial einer österreichischen Gesellschaft (1220 Schäden), das er einer Arbeit von

Lethay [7] entnahm, bestätigt Myslivec die Richtigkeit seiner Annahme. Aus 20 Schadengruppen mit gleichen Intervallbreiten von je 400 Schilling ermittelt er die empirischen Häufigkeiten und bestimmt daraus auch die Konstanten a und b der Pareto-Funktion mit Hilfe der Methode der kleinsten Quadrate. Durch einen graphischen Vergleich der theoretischen Geradengleichung

$$\log g(y) = \log a - b \log y \qquad (1a)$$

(theoretische Pareto-Funktion in logarithmischer Transformation)

mit dem empirischen Punktediagramm, das durch Auftragen der Logarithmen der Schadengrössen und der entsprechenden beobachteten Häufigkeiten in ein Koordinatennetz mit zwei logarithmischen Skalen gebildet wurde, erzielt er in der Tat eine recht gute Übereinstimmung, vor allem bei den mittelgrossen Schäden (2000–6000 Schilling); die grossen Schäden über 8000 Schilling wurden wegen ihres seltenen Auftretens nicht in die Rechnung einbezogen, ansonst das Bild der Linearität gestört worden wäre.

Die Übernahme dieser mathematisch einfach gebauten Pareto-Funktion für unsere Zwecke musste aber verworfen werden aus folgenden Gründen:

1. Der *Verteilungscharakter der häufigen kleinen Schäden* ist durch die grosse Intervallbreite verwischt; die Wahl einer solchen Klassenbreite ist aber hauptsächlich durch das analytische Verhalten der Funktion (1) für kleine y bedingt (für $y \to o$ strebt $g(y) \to \infty$), womit auch die unbefriedigende Anpassung bei den kleinen Schäden erklärlich ist.
2. Die über den ganzen Definitionsbereich geforderte Linearität gemäss (1a) ist nicht erfüllt wie in einer französischen Untersuchung aus dem Jahre 1945 anhand eines viel umfangreicheren Beobachtungsmaterials als desjenigen, das Myslivec zur Verfügung stand, nachgewiesen wurde (Pellegrin [5]); in Wirklichkeit ergibt sich für die Gesamtheit der Sach- und Personenschäden eher ein gegen die Abszissenachse hin konkaves Punktediagramm, das eventuell durch Bildung einer Gruppe von drei sich schneidenden Geraden ausgeglichen werden könnte; Pellegrin verwirft deshalb den Pareto-Ansatz vom theoretischen Standpunkt aus.

3. Es ergäbe sich demnach ein viel zu ungenauer theoretischer Durchschnitt M^1, da eine einzige, nur durch die mittelgrossen Schäden gemäss (1a) bestimmte Gerade die bedeutenden Schäden weit überschätzen würde.

Die Pareto-Verteilung kann ihrer Einfachheit wegen gute Dienste zur rationellen und approximativen Berechnung von Versicherungssätzen in Abhängigkeit von der Höhe der maximalen Haftungsgrenze leisten, wie es Myslivec gezeigt hat. Für die Bestimmung eines zuverlässigen durchschnittlichen Schadenbetrages aber ist die Annahme der einfachen Pareto-Funktion als Grundgesamtheit nicht empfehlenswert.

b) Riebesell [6] schlägt für verschiedene Versicherungszweige folgende Ausgleichsfunktion vor:

$$g(y) = ae^{-by} \tag{2}$$

c) Lethay [7] erbringt in seiner Arbeit den Beweis, dass diese von Riebesell empfohlene Funktion für die praktische Anwendung in der Autohaftpflichtversicherung nicht geeignet ist und stellt selbst eine andere Funktion auf:

$$g(y) = b^{c^y} - d \tag{3}$$

Nach Myslivec [4] bewährt sich auch diese Formel nicht bei der praktischen Auswertung, da die Berechnung ihrer Konstanten recht schwierig und kompliziert ist.

d) Everts [8] veröffentlichte umfangreiches Zahlenmaterial über Autohaftpflicht-Schäden aus den Jahren 1925–1931, das sich auf eine Gemeinschaftsstatistik von mehreren deutschen Haftpflicht-Versicherungsgesellschaften stützt. Um die Bedarfsprämien und die Schadenreserven bestimmen zu können, bedient er sich der theoretischen Verteilungsfunktion:

$$G(y) = \frac{a\,y^k}{y^k + b} \tag{4}$$

y bedeutet hier die jeweilige Schadengrenze, G(y) die Zahl der entschädigten Fälle.

Er erzielte dadurch eine recht befriedigende Übereinstimmung mit der beobachteten Summenfunktion. Die für die Statistik benutzten

Schadengrenzen sind willkürlich und nicht äquidistant festgelegt. Die der Verteilungsfunktion $G(y)$ zugeordnete theoretische Häufigkeitsfunktion $g(y)$ würde man durch Ableiten des Ausdruckes (4) gewinnen, doch ist dieses Vorgehen für unsere Zwecke des mathematisch komplizierten Aufbaues von $G(y)$ wegen nicht angezeigt.

e) Konac [9] drückt die Häufigkeiten in den einzelnen Grössengruppen durch *Pearson'sche Kurven* aus, indem er zunächst mittels der Momentenmethode den entsprechenden Typ der Funktion und ihre Konstanten bestimmt. Für die Gesamtschadenverteilung einer Gruppe von N Versicherungen der gleichen Beobachtungsmasse empfiehlt er Typus I und IV, deren Konstanten aber auf sehr umständliche Weise aus den 4 ersten Momenten der empirischen Verteilung bestimmt werden müssen. Um einigermassen zuverlässige Werte zu erhalten, ist sehr umfangreiches Beobachtungsmaterial nötig.

Alle unter a bis e aufgeführten Ansätze fallen für unsere Untersuchungen ausser Betracht. Sie geben entweder die wirklichen Verhältnisse zu wenig gut wieder ((1) und (2)), oder deren mathematische Gestalt ist zu kompliziert besonders zur praktischen Bestimmung der Parameter ((3), (4), e); sie sind für eine gute und zuverlässige indirekte Durchschnittsbestimmung ungeeignet, da zusätzlich gerade die komplizierten Häufigkeitsfunktionen neueren, statistischen Bewertungsmethoden schwer zugänglich sind. Zu Ausgleichszwecken dürften sie alle mehr oder weniger taugen, besonders wenn das anzupassende Material sich aus Erfahrungen über einen längeren Zeitraum zusammensetzt, was aber wiederum unserer Zielsetzung widerspricht.

Eine willkürlich gewählte Klasseneinteilung der Schadengrössen ist übrigens auch vom mathematischen Standpunkt aus unbefriedigend, hauptsächlich beim Grenzübergang von der Stichprobenverteilung zur kontinuierlichen Grundgesamtheit. Deshalb werden im Folgenden die Hypothesen aufgestellt, dass die *logarithmisch transformierte Merkmalsvariable* $x = \log y$ entweder normalverteilt (12) oder gemäss *Pearson Typ III* verteilt ist (13), und anhand unseres Beobachtungsmaterials werden diese Annahmen geprüft. Durch diese Transformation ergibt sich eine geometrisch progressiv wachsende Merkmalsskala für die Zufallsvariable y, was sowohl die praktische als auch die theoretische Behandlung unseres Problems erleichtert.

12 DIE LOGNORMALE VERTEILUNG

121 Hypothese

Die *transformierte Zufallsvariable* $x = \ln y$ ist *normalverteilt:* (ln: natürlicher Logarithmus zur Basis e, der hier zur Erleichterung der theoretischen Betrachtungen benützt wird. Die gewonnenen Resultate können dann ohne weiteres auf die in der praktischen Auswertung gebrauchten 10er-Logarithmen übertragen werden)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\xi}{\sigma}\right)^2} \tag{1}$$

Nach den allgemeinen Transformationsprinzipien von Zufallsvariablen (siehe z. B. KENDALL [A], Bd. I, S. 16) gilt dann für die *Häufigkeitsfunktion der ursprünglichen zufälligen Variablen* y = Schadenbetrag:

$$g(y) = \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{y} e^{-\frac{1}{2}\left(\frac{\ln y-\xi}{\sigma}\right)^2} \tag{2}$$

Dies ist die zweiparametrige *logarithmische Normalverteilung*, von nun an kurz „*Lognormalverteilung*" genannt.

Bekanntlich hängen die *transformierte und die ursprüngliche Häufigkeitsfunktion* durch die Differentialgleichung zusammen:

$$f(x)\,dx = g(y)\,dy \tag{3}$$

122 Allgemeines über die Lognormalverteilung und ihr Anwendungsgebiet; Literaturhinweise

Die Lognormalverteilung in ihrer zwei- oder dreiparametrigen Gestalt ist schiefsymmetrisch; sie eignet sich gut als theoretisches Modell zur Darstellung von biologischen, wirtschaftlichen, demographischen und astronomischen Grundgesamtheiten. Bei der im Gegensatz zu unserer Form (2) drei Parameter aufweisenden Lognormalverteilung ist die neu hinzukommende Konstante für die Lage des Ursprungs verantwortlich; in unserem Fall ist diese aber a priori durch die spezielle Problemstellung bestimmt (y = o).

Die Abhandlungen über diese vielseitige Häufigkeitsfunktion sind recht zahlreich; aus der Fülle seien nur ein paar grundlegende Arbeiten neueren Datums herausgegriffen: PAE-TSI-YUAN [10], FINNEY [11],

Cohen [12] und Gebelein [13], der ihr eine so grosse Bedeutung zumisst, dass er sie geradezu als Normalverteilung 2. Art bezeichnet. Für weitere Einzelheiten sei auf das reichhaltige, bis 1936 ziemlich vollständige Literaturverzeichnis in der Schrift von d'Addario [14] auf Seite 106 unten verwiesen.

Es waren vor allem die italienischen Statistiker Amoroso [15] und d'Addario [16], die die auf Edgeworth zurückgehende Idee der logarithmischen Transformation der Merkmalsvariablen zur Normalisierung aufnahmen und sie in der *Unfall- und Haftpflichtversicherung* mit Erfolg anwandten, insbesondere zur Ausgleichung von Renten- und Schadenverteilungen. Auch die französischen Aktuare Risser [17], Henry [18] und Pellegrin [5] machten sich diese „Edgeworth-Transformation" in ihren Untersuchungen zunutze, die darin besteht, dass die transformierte zufällige Variable

$$z = a \log y + b$$

einer standardisierten Normalverteilung unterliegt (Durchschnitt o, Streuung 1):

$$h(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

Praktisch wurden die Konstanten a und b aus den beobachteten Werten nach der Methode der kleinsten Quadrate bestimmt. Unsere in 121 ausgedrückte Hypothese ist hier mathematisch etwas anders formuliert, doch ihrem Wesen nach entspricht sie genau der obigen Grundidee.

Wir werden uns im Folgenden an die Formen (1) und (2) halten, da dadurch die Genesis der Lognormalverteilung aus der gewöhnlichen Normalverteilung und ihre gegenseitigen Zusammenhänge besser zum Ausdruck kommen.

123 Die wichtigsten theoretischen Kenngrössen der Lognormalverteilung; Zusammenhänge mit der Gauss'schen Normalverteilung

Die folgenden Formeln können alle leicht mit Hilfe der Grundbeziehungen (1)–(3) hergeleitet werden, oder sie finden sich in jeder Abhandlung über die Lognormalverteilung.

a) *Momente der Häufigkeitsfunktion* (12.2):

$$\mu'_\nu = \int_0^\infty y^\nu g(y)\,dy = \int_{-\infty}^{\infty} e^{\nu x} f(x)\,dx = E[y^\nu]$$

$$\mu'_\nu = e^{\nu\xi + \frac{\nu^2\sigma^2}{2}} \tag{4}$$

daraus folgt für den *Durchschnitt* $\mu(y) = \mu'_1(y)$

$$\mu(y) = e^{\xi + \frac{1}{2}\sigma^2} \tag{5}$$

und für die *Streuung* $\mu_2(y) = \mu'_2 - \mu^2(y) = E[y - E(y)]^2$

$$\mu_2(y) = \sigma^2(y) = e^{2\xi + \sigma^2}(e^{\sigma^2} - 1) = \mu^2(y)[e^{\sigma^2} - 1] \tag{6}$$

Schreiben wir zur Abkürzung
$\tau = e^{\sigma^2}$, so wird die *mittlere quadratische Abweichung:*

$$\sigma(y) = \mu(y)[\tau - 1]^{\frac{1}{2}} \tag{7}$$

Schiefe γ_1 und *Exzess* γ_2 seien der Vollständigkeit halber aufgeführt:

$$\left.\begin{aligned} \gamma_1 &= (\tau - 1)^{\frac{1}{2}}(\tau + 2) \\ \gamma_2 &= (\tau - 1)(\tau^3 + 3\tau^2 + 6\tau + 6) \end{aligned}\right\} \tag{8}$$

Je grösser σ^2, desto stärker weicht die Lognormalverteilung von der gewöhnlichen Normalverteilung ab, und desto schiefer ist sie.

b) Das *geometrische Mittel* η_G und der *Zentralwert* η_Z fallen zusammen. *Ihnen entspricht* in der transformierten (Normal-)Verteilung $f(x)$ *der Durchschnitt* ξ

$$\eta_G = \eta_Z = e^\xi \tag{9}$$

denn: $\xi = \int_{-\infty}^{\infty} x f(x)\,dx = \int_0^\infty \ln y\, g(y)\,dy = \ln \eta_G$ nach Def.

und: $\frac{1}{2} = \int_{-\infty}^{\xi} f(x)\,dx = \int_0^{y=e^\xi} g(y)\,dy = \int_0^{\eta G} g(y)\,dy = \int_0^{\eta Z} g(y)\,dy$

Das *geometrische Mittel* η_G spielt also in der *Lognormalverteilung* die gleich wichtige Rolle *wie* der *Durchschnitt* ξ in der *gewöhnlichen Normalverteilung.*

Den *dichtesten Wert* η_D erhalten wir durch Nullsetzen der Ableitung von g(y) oder direkt, wenn wir g(y) in die Form einer Gauss'schen Normalverteilung bringen:

aus (2): $$g(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}[(\ln y - \xi)^2 + 2\sigma^2 \ln y]}$$

Durch quadratische Ergänzung im Exponenten folgt:

$$g(y) = \frac{e^{\frac{\sigma^2}{2} - \xi}}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(\ln y - \xi + \sigma^2)^2} \tag{2a}$$

und daraus: $$\eta_D = e^{\xi - \sigma^2} \tag{10}$$

Durch einen Vergleich von (5), (9) und (10) erhalten wir die Ungleichung: $$\eta_D < \eta_G = \eta_Z < \mu(y) \tag{11}$$

c) Der *mittleren quadratischen Abweichung* σ entspricht in der Lognormalverteilung ein „*Abweichungsverhältnis* $\sigma_L(y)$" *von* η_G.

Wenn wir $\sigma_L(y)$ *wie folgt definieren:*

$$\ln^2 \sigma_L(y) = \int_0^\infty (\ln y - \ln \eta_G)^2 \, g(y)\, dy = \int_0^\infty \left(\ln \frac{y}{\eta_G}\right)^2 g(y)\, dy$$

wird: $$\sigma^2 = \int_{-\infty}^{\infty} (x - \xi)^2 f(x)\, dx = \int_0^\infty (\ln y - \ln \eta_G)^2 \, g(y)\, dy = \ln^2 \sigma_L(y)$$

$$\rightarrow \sigma_L(y) = e^{\sigma} \quad (\sigma_L(y) \geq 1) \tag{12}$$

Wahrscheinlichkeiten von Abweichungen von ξ bezw. η_G können dann wie folgt miteinander in Beziehung gesetzt und aus einer Tafel der Gauss'schen Fehlerfunktion bestimmt werden:

$$\int_{\xi + z_1\sigma}^{\xi + z_2\sigma} f(x)\, dx = \int_{\eta_G \cdot \sigma_L^{z_1}}^{\eta_G \cdot \sigma_L^{z_2}} g(y)\, dy = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-\frac{\varphi^2}{2}} d\varphi \quad (z_1 < z_2)$$

da $z = \frac{x - \xi}{\sigma} = \frac{\ln y - \ln \eta_G}{\ln \sigma_L(y)}$ standardisiert normalverteilt (0,1) ist.

Anschauliche Interpretation für $z_1 = -z$, $z_2 = z$ $(z = 1, 2, \ldots)$

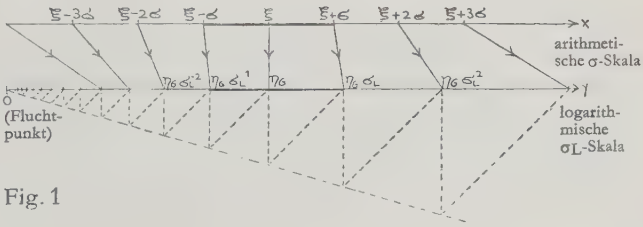


Fig. 1

124 Bestimmung der Parameter ξ und σ:

Die Bestimmung der unbekannten Parameter ξ und σ geschieht natürlich am besten aus den transformierten Stichprobengrössen $x = \ln y$ durch die „erschöpfenden" („sufficient", siehe 134) Schätzwerte $\bar{x}$ = *Stichprobendurchschnitt* und s^2 = *Stichprobenstreuung:*

$$\xi = \bar{x} = \frac{1}{n} \sum_1^n x_j \tag{13}$$

$$\sigma = s(x) = \sqrt{\frac{1}{n-1} \sum_1^n (x_j - \bar{x})^2} \tag{14}$$

Auch hier bestimme ich *durch* $\bar{x}$ den *Logarithmus des geometrischen Mittels* y_G der Stichprobe (y) und durch $s(x)$ *den Logarithmus des Stichproben-Abweichungsverhältnisses* $s_L(y)$

denn: $$\ln y_G = \frac{1}{n} \sum_1^n \ln y_j = \ln \left[\left(\prod_1^n y_j \right)^{\frac{1}{n}} \right]$$

nach Definition des geometrischen Mittels. *Definieren wir* wie oben mit:

$$\ln s_L(y) = \sqrt{\frac{1}{n-1} \sum_{j=1}^n (\ln y_j - \ln y_G)^2} = \sqrt{\frac{1}{n-1} \sum_j \left(\ln \frac{y_j}{y_G} \right)^2}$$

den *Logarithmus des Abweichungsverhältnisses* $s_L(y)$ der Stichprobe, so wird:

$$\bar{x} = \ln y_G; \quad y_G = e^{\bar{x}} \tag{15}$$

$$s(x) = \ln s_L(y); \quad s_L(y) = e^{s(x)} \tag{16}$$

y_G und $s_L(y)$ bilden ersichtlich „erschöpfende" Schätzwerte für η_G und $\sigma_L(y)$ der y-Grundgesamtheit.

125 Schätzwerte für den Durchschnitt $\mu(y)$ und die Streuung $\sigma^2(y)$

Finney [11] leitet die *„wirksamsten" Schätzwerte* für die unbekannten $\mu(y)$ und $\sigma^2(y)$ der Originalgrundgesamtheit $g(y)$ aus den transformierten, normalverteilten Stichprobendaten x her. Von mehreren, nach verschiedenen Methoden gewonnenen Schätzwerten $t_1, \ldots t_k$ für ein- und denselben Parameter θ heisst nach der von R. A. Fisher [E] begründeten Schätztheorie derjenige der *„wirksamste"* (t_w) („most efficient estimate"), der die *kleinste Stichprobenstreuung* $\sigma^2(t_w)$ aufweist.

Ein Schätzwert ist als Funktion der n zufälligen Variablen x selbst eine Zufallsvariable mit einer bestimmten Häufigkeitsfunktion $f(t_1)$. Die Stichprobenstreuung von t_1 ist also durch

$$\sigma^2(t_1) = \int_{-\infty}^{\infty} (t_1 - \theta)^2 f(t_1) dt_1 = E[t_1 - \theta]^2 \text{ gegeben.}$$

Durch den Begriff der „*Wirksamkeit*“ *(efficiency)* $W(t_i)$ irgendeines Schätzwertes t_i:

$$W(t_1) = \frac{\sigma^2(t_w)}{\sigma^2(t_i)} \text{ (meistens in Prozenten ausgedrückt)}$$

kann somit die Güte eines nach irgendeiner Methode bestimmten t_i beurteilt werden, falls t_w und $\sigma^2(t_w)$ bekannt sind.

Die von Finney [11] gewonnenen wirksamsten Schätzwerte m und v für $\mu(y)$ und $\mu_2(y)$ lauten:

$$m = e^{\bar{x} + \frac{1}{2}s^2} \left\{1 - \frac{s^2(s^2+2)}{4n} + 0\left(\frac{1}{n^2}\right)\right\} \tag{17}$$

$$v = e^{2\bar{x}+s^2} \left\{e^{s^2}\left[1 - \frac{2s^2(2s^2+1)}{n} + 0\left(\frac{1}{n^2}\right)\right] - \left[1 - \frac{s^2(s^2+2)}{n} + 0\left(\frac{1}{n^2}\right)\right]\right\} \tag{18}$$

wobei n = *Stichprobenumfang*, $\bar{x}$ und s^2 durch (13) und (14) definiert sind.

Durch einen Vergleich mit (5) und (6) erkennen wir das Bildungsgesetz; anstelle der unbekannten Parameter ξ und σ^2 sind ihre besten Schätzwerte $\bar{x}$ und s^2 einzusetzen, die Korrekturglieder in den geschweiften Klammern werden für grössere n und kleinere s^2 rasch unbedeutend.

Finney bestimmte noch die Wirksamkeiten $W(\bar{y})$ und $W[s^2(y)]$ des direkt geschätzten Durchschnittes $\bar{y} = \frac{1}{n}\sum_{j=1}^{n} y_j$ und der direkten Streuung $s^2(y) = \frac{1}{n-1}\sum_{1}^{n}(y_j - \bar{y})^2$, die auch Schätzwerte von $\mu(y)$ und $\sigma^2(y)$ sind (Momentenmethode). Die recht langwierigen Ableitungen und die komplizierten Ausdrücke seien hier unterlassen; man schliesst, dass $W(\bar{y})$ für ein $\sigma^2 < 0{,}69$ noch über 80% ist, d.h. der *direkte Durchschnitt* $\bar{y}$ ist eine recht *gute Schätzung von* $\mu(y)$; die Wirksamkeit $W[s^2(y)]$ aber wird sehr rasch klein, für $\sigma^2 = 0{,}69$ beträgt sie nur noch 28%. $s^2(y)$ *ist also unwirksam zur Schätzung von* $\sigma^2(y)$.

126 Anwendung. Überprüfung der Hypothese von 121 anhand unseres Beobachtungsmaterials

Aus dem uns zur Verfügung stehenden Erfahrungsmaterial wurden die *Autohaftpflicht-Schäden der „Berner Allgemeinen"* aus dem Jahre 1950 herangezogen; die Wahl dieses Jahres ist rein willkürlich erfolgt.

Passende Wahl der logarithmischen Merkmalsskala:

Um vernünftige Besetzungszahlen (Häufigkeiten f_j) zu erhalten, wurde eine *äquidistante Argumentenfolge* $x = \log_{10} y$ mit einer Klassenbreite von

$$\Delta = \tfrac{1}{6} = 0{,}16\,667$$

so angesetzt, dass ganzzahlige $x = 1, 2, 3, 4, 5$ ($y = 10, 100, 1000, 10000, 100000$ Fr.) in eine *Klassenmitte* x_j zu liegen kommen; die *Klassengrenzen* x_k/x_{k+1} bauen sich also aus den Klassenmitten x_j wie folgt auf:

$$x_k = x_j - \frac{\Delta}{2}\,;\; x_{k+1} = x_j + \frac{\Delta}{2} \quad (j = 1, 2, \ldots, l;\; k = 1, 2, \ldots l):$$

$$\ldots, 1 \pm \tfrac{1}{12}, 1\tfrac{1}{6} \pm \tfrac{1}{12}, 1\tfrac{2}{6} \pm \tfrac{1}{12}, \ldots, 2 \pm \tfrac{1}{12}, \ldots \text{ etc.}$$

(siehe Anhangtabelle 1, Kolonnen (1) und (2)).

Diese Festsetzung der Klassen ist natürlich willkürlich, doch hat sie sich vom praktischen Standpunkt aus als sinnvoll erwiesen, liegen doch die häufig vorkommenden Schäden von 10, 100 und 1000 Fr. in der Mitte eines Intervalls.

Für die *geometrisch progressiv wachsende Argumentenfolge* y ergibt sich demnach ein dem $\Delta = \frac{1}{6}$ entsprechendes konstantes *Verhältnis* von

$$q = \text{antilog}\, \tfrac{1}{6} = 1{,}4678$$

für aufeinanderfolgende *Klassenmitten* y_j und *Klassengrenzen* y_k/y_{k+1}:

$$y_{j+1} = q y_j \quad (j = 1, 2, \ldots l)$$
$$y_{k+1} = q y_k \quad (k = 1, 2, \ldots l)$$

y_k: $\ldots,$ $\underbrace{8{,}25}_{10.q^{-\frac{1}{2}}}$ — $\underbrace{12{,}11}_{10.q^{\frac{1}{2}}}$, 12,12 — 17,77, $\ldots\ldots$, 82,54 — 121,1, $\ldots$

y_j: $\ldots,$ 10 , 14,68 , $\ldots\ldots$, 100 , $\ldots$

(siehe Anhangtabelle 1, Kolonnen (3) und (4)).

Die gesamte Variationsbreite der Schäden von rund 3–120000 Fr. wurde dadurch in l = 28 Klassen unterteilt.

Die *empirischen Häufigkeiten* $f_j (j = 1, 2, \ldots l)$ wurden dem Schadenregister „Motorfahrzeughaftpflicht-Schäden 1950" nach Aussonderung der durch Lastwagen und Motorräder verursachten Schäden entnommen; da kein Lochkartenmaterial zur Verfügung stand, mussten die einzelnen Schadenbeträge von Hand in die entsprechenden Klassen y_k / y_{k+1} eingetragen werden; es wurden dabei noch Jahresuntergruppen von je 200 Schäden gebildet, die uns dann im 3. Kapitel für stichprobentheoretische Untersuchungen dienen werden. 1950 waren *1297 Schadenfälle* registriert, von denen aber *329 ohne Folgen* blieben und für unsere Berechnungen ausser Betracht fallen; die Zahl der Schäden beträgt demnach $n^{1950} = 968$; deren Verteilung nach der Höhe ihres Schadenbetrages (f_j^{1950}) ist der Kolonne (5) der Anhangtabelle 1 zu entnehmen, sowie dem Staffelbild in Fig. 2. Bei den noch *nicht endgültig erledigten Fällen* (1950 waren es 8, zur Hauptsache Rentenfälle) wurde als Schadensumme die bereits *getätigten Zahlungen + Schadenreserve Ende 1950* eingesetzt, also Schätzwerte, die aber den Allgemeincharakter der Schadenverteilung nicht allzu stark beeinflussen dürften.

Wir haben nun die *Hypothese 121 zu prüfen*, ob $x = \ln y$ *normalverteilt* liegt. Dazu wurde vorerst eine *Ausgleichung mit Hilfe des Ansatzes (1)* vorgenommen, wobei aus Zweckmässigkeitsgründen anstelle der natürlichen Logarithmen die 10er-Logarithmen verwendet wurden:

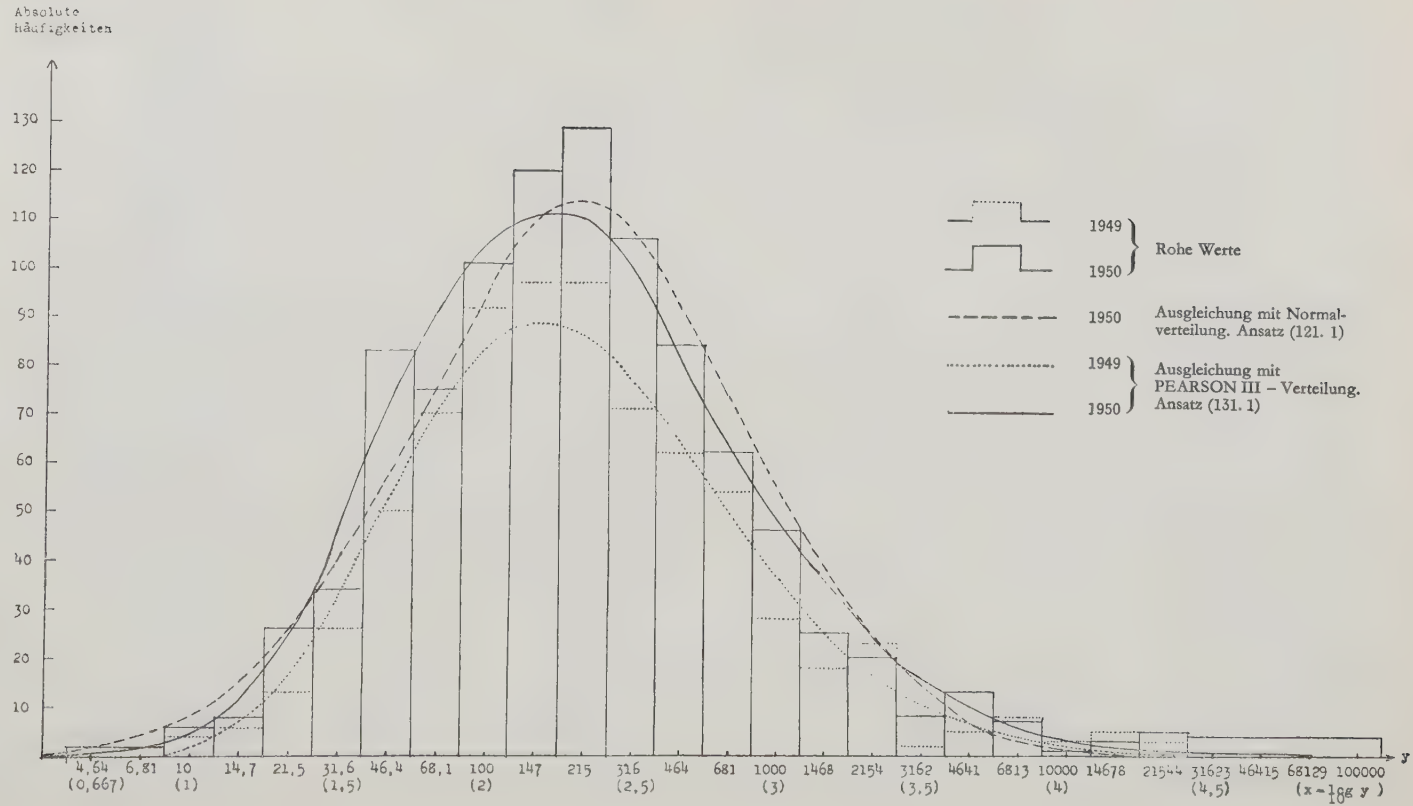
$$x' = \log_{10} y = \ln y \,.\, \log e = x \,.\, \log e$$

Die *Parameter* ξ und σ bestimmen sich sinngemäss aus (13) und (14) und ergaben im vorliegenden Fall folgende Werte:

$$\xi^* = \bar{x}' = 2{,}32333 = \log_{10} y_G^{1950}$$
$$\sigma^{*2} = s'^2(x) = 0{,}34869;\ \sigma^* = s'(x) = 0{,}5905$$

Die Bestimmung der *theoretischen Häufigkeiten* φ_j erfolgte nun nach dem üblichen Verfahren unter Verwendung einer Tabelle für Gauss'sche Verteilungsfunktion durch Bildung der standardisierten Abweichungen $t_k = \frac{x'_k - \xi^*}{\sigma^*}$ und Bestimmung der zugehörigen theoretischen

Figur 2


Absolute Häufigkeiten
130
120
110
100
90
80
70
60
50
40
30
20
10
4,64 6,81 10 14,7 21,5 31,6 46,4 68,1 100 147 215 316 464 681 1000 1468 2154 3162 4641 6813 10000 14678 21544 31623 46415 68129 100000
(0,667) (1) (1,5) (2) (2,5) (3) (3,5) (4) (4,5)
(x ~ lg y)
y
1949
1950
Rohe Werte
1950
Ausgleichung mit Normalverteilung. Ansatz (121. 1)
1949
1950
Ausgleichung mit PEARSON III – Verteilung. Ansatz (131. 1)

Klassenhäufigkeiten φ_j. Das *Resultat dieser Ausgleichung* ist in der Anhangtabelle 1, Kolonne (6) und durch die gestrichelte Kurve in der Fig. 2 wiedergegeben. Wir erkennen schon hier die schlechte Anpassung der grossen Schäden an die theoretische Normalverteilung.

Um beurteilen zu können, ob die Abweichungen zwischen beobachteter und theoretischer Verteilung zufällig oder systematisch sind, bedienten wir uns des bekannten χ^2-Testes (siehe Anhangtabelle 1): $\chi^2 = 58{,}7$ mit $N = 17$ Freiheitsgraden (19 berücksichtigte Gruppen minus Zahl der zu bestimmenden Parameter = 2); die *Abweichungen reichen weit über das zulässige Zufallsmass hinaus:* $P(\chi^2) < 0{,}001$. Das schlechte Resultat ist ersichtlich der letzten stark zusammengefassten Gruppe der grossen Schäden zuzuschreiben.

Wir vermuten nun auch sofort, dass der *theoretische Durchschnitt* m_y^{1950} *nach* FINNEY *gemäss (17)* gegenüber dem direkten Durchschnitt $\bar{y}^{1950}$ viel zu klein ausfallen wird, was die Rechnung bestätigt:

Nach (17) wird in 1. Näherung:

$$\ln m_y^{1950} = \bar{x} + \tfrac{1}{2}s^2 = \ln 10 \,.\, \log m_y^{1950}$$

$$\bar{x} = \ln y_G = \ln 10 \,.\, \log y_G = \bar{x}' \ln 10$$

$$s^2(x) = \frac{1}{n-1} \sum_{j=1}^{i} f_j (\ln y_j - \ln y_G)^2 = s'^2(x) \,.\, \ln^2 10$$

Mit den obigen Werten $\bar{x}' = 2{,}32333$ und $s'^2(x) = 0{,}34869$, sowie $\ln 10 = 2{,}302585$ wird:

$$\log m_y^{1950} = \bar{x}' + \frac{s'^2}{2} \cdot \ln 10 = 2{,}72477$$

$$m_y^{1950} = 530{,}6$$

Der direkte Durchschnitt $\bar{y}^{1950}$ ergibt nach Anhangtabelle 1:

$$\bar{y}^{1950} = 856{,}8$$

Folgerung: Es dürfte nun erwiesen sein, dass die *Hypothese der Lognormalverteilung als Häufigkeitsfunktion für die Gesamtheit der Körper- und Sachschäden verworfen werden* muss, da durch sie die wichtigen grossen Schäden viel zu wenig berücksichtigt werden. Die Ausgleichung mit (1) könnte befriedigen, wenn nur die Sachschäden analysiert würden,

da vermutlich die grossen Körperschäden eine Verteilung für sich bilden, voraussichtlich eine Poisson-Verteilung. Eine solche Trennung in Körper- und Sachschäden war aber an unserem Material nicht möglich.

Die Vermutung liegt nahe, dass eine asymmetrische Glockenkurve mit positiver Schiefe die Anpassung an die empirische Verteilung stark verbessern könnte. Auf der Suche nach einer solchen theoretischen Häufigkeitsverteilung fiel die Wahl bald auf die Pearson-Verteilung Typ III, von der im nächsten Abschnitt die Rede sein wird.

13 DIE PEARSON TYP III-VERTEILUNG

131 Allgemeines:

Die Pearson-Verteilung vom Typus III findet sich in der statistischen Literatur in verschiedenen Darstellungen und zwar je nach der Lage des Ursprungs in zwei- oder dreiparametriger Gestalt. Eine gute Übersicht über die dreiparametrige Form, wie sie hauptsächlich in der Ausgleichungsrechnung gebraucht wird, ist in Elderton [19] wiedergegeben (Ursprung im dichtesten Wert oder im Durchschnitt).

Für unsere Studie verwenden wir die zweiparametrige Form mit dem Ursprung in Null:

$$f(x) = \frac{\alpha^\gamma}{\Gamma(\gamma)} e^{-\alpha x} x^{\gamma-1} \quad \begin{matrix} (0 < x < \infty) \\ (\alpha > 0, \gamma > 0) \end{matrix} \qquad (1)$$

In dieser oder ähnlicher Gestalt treffen wir sie auch meistens in der mathematisch-statistischen Literatur an (Kendall [A], Cramer [B] u.a.); in spezieller Form $\left(\alpha = \frac{1}{2}, \gamma = \frac{n}{2}\right)$ spielt sie in der Stichprobentheorie unter dem Namen „χ^2-*Verteilung*" eine fundamentale Rolle.

132 Hypothese:

Die transformierte Zufallsvariable $x = \log y$ ist *gemäss* (1) verteilt. Analog wie bei (12. 1) folgt wiederum aus

$$g(y)\,dy = f(x)\,dx$$

für die *theoretische Häufigkeitsfunktion des Schadenbetrages* y:

$$g(y) = \frac{(\alpha \log e)^{\gamma}}{\Gamma(\gamma)} y^{-\alpha \log e - 1} (\ln y)^{\gamma - 1} (1 < y < \infty) \qquad (2)$$

die wir „*logarithmische* PEARSON *III-Verteilung*" nennen wollen.

133 Theoretische Kenngrössen der Verteilung (1):

charakteristische Funktion $\varphi(t) = \int_0^{\infty} e^{itx} f(x)dx$: nach KENDALL [A], Band I, Seite 55ff.:

$$\varphi(t) = \left(1 - \frac{it}{\alpha}\right)^{-\gamma} (i = \text{imaginäre Einheit}) \qquad (3)$$

Momente $\mu'_{\nu}(x)$ und Kumulanten $\varkappa_{\nu}(x)$:

$$\mu'_{\nu}(x) = \frac{\gamma(\gamma+1)\ldots(\gamma+\nu-1)}{\alpha^{\nu}} \qquad (4)$$

$$\varkappa_{\nu}(x) = \frac{\gamma(\nu-1)!}{\alpha^{\nu}} \qquad (5)$$

Daraus gewinnen wir leicht *Durchschnitt* $\mu(x) = \mu_1'(x) = \varkappa_1$
Streuung $\mu_2(x) = \varkappa_2 = \mu_2' - \mu^2(x)$, *Schiefe* $\gamma_1 = \frac{\mu_3}{\mu_2^{3/2}} = \frac{k_3}{k_2^{3/2}}$ und
Exzess $\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 = \frac{k_4}{k_2^2}$

$$\mu(x) = \frac{\gamma}{\alpha} \qquad (6)$$

$$\mu_2(x) = \sigma^2(x) = \frac{\gamma}{\alpha^2} \qquad (7)$$

$$\gamma_1 = \frac{2}{\sqrt{\gamma}} \qquad (8)$$

$$\gamma_2 = \frac{6}{\gamma} \qquad (9)$$

Für grosse γ strebt demnach die PEARSON III-Verteilung gegen die Gauss'sche Normalverteilung.

a) nach dem Prinzip des Vergleichs der Momente (Momentenmethode):

Setzen wir in (6) und (7) für die theoretischen Momente $\mu(x)$ und $\mu_2(x)$ ihre entsprechenden empirischen Momente $\bar{x}$ und $s^2(x)$, so gewinnen wir die 2 Bedingungsgleichungen:

$$\bar{x} = \frac{\gamma}{\alpha}$$

$$s^2(x) = \frac{\gamma}{\alpha^2}$$

Durch Auflösen nach γ und α erhalten wir die „*Momenten*"-*Schätzwerte* α^* und γ^*:

$$\gamma^* = \frac{\bar{x}^2}{s^2(x)} \tag{10}$$

$$\alpha^* = \frac{\bar{x}}{s^2(x)} = \frac{\gamma^*}{\bar{x}} \tag{11}$$

b) nach dem Prinzip der wahrscheinlichsten Parameter (Maximum-Likelihoodmethode):

Diese von R. A. Fisher [E] in seiner schon erwähnten Schätztheorie entwickelte Methode gibt gewissermassen beste Schätzwerte $\hat{\theta}_l$ für l unbekannte Parameter θ_l einer analytischen Grundgesamtheit $f(x; \theta_1, \theta_2, \ldots, \theta_l)$. Nach dieser Methode werden die Parameter $\theta_l (i = 1, 2, \ldots l)$ so bestimmt, dass die *Wahrscheinlichkeit* $L(x, \theta_l)$ *für das Auftreten der beobachteten Konstellation der Stichprobengrössen* $x_1, x_2, \ldots . x_n$ *ein Maximum erreicht:*

$$L(x_1, \ldots, x_n; \theta_l) = \prod_{k=1}^{n} f(x_k, \theta_l) = \text{Maximum} \tag{A}$$

wegen (A) muss dann notwendigerweise gelten:

$$\frac{\partial L}{\partial \theta_1} = 0, \frac{\partial L}{\partial \theta_2} = 0, \ldots ., \frac{\partial L}{\partial \theta_l} = 0$$

Meistens wird die zur Berechnung handlichere logarithmische Form gebraucht:

$$\frac{\partial \ln L}{\partial \theta_1} = 0, \frac{\partial \ln L}{\partial \theta_2} = 0, \ldots ., \frac{\partial \ln L}{\partial \theta_l} = 0 \tag{B}$$

$$\left(\text{aus } \frac{\partial L}{\partial \theta} = 0 \rightarrow \frac{\partial \ln L}{\partial \theta} = \frac{\partial L}{\partial \theta} \frac{1}{L} = 0\right)$$

Wenn Lösungen $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_l$ der l Bestimmungsgleichungen (B) existieren, so bilden diese nach der allgemeinen Theorie (siehe z. B. KENDALL [A], Band II, Kapitel 17) deshalb *optimale Schätzwerte* von $\theta_1, \theta_2, \ldots, \theta_l$, weil sie folgende Eigenschaften aufweisen:

1. Sie sind *eindeutig.*
2. Das *Verteilungsgesetz der geschätzten Parameter strebt für grosse* n *gegen die Normalverteilung.*
3. Sie sind *„passend"* („consistent"). Allgemein heisst ein Schätzwert t „passend", falls die aus der Häufigkeitsfunktion f(t) des Parameters t zu entnehmende Wahrscheinlichkeit, dass der berechnete Wert von t um mehr als einen beliebig kleinen Betrag ε vom wahren Wert θ abweicht, bei wachsendem Stichprobenumfang beliebig nahe gegen Null sinkt.
4. Sie sind für *grosse Stichproben am „wirksamsten"* („most efficient". Begriff siehe vorn unter 125).
5. Sie sind *„erschöpfend"* („sufficient"), *falls überhaupt* solche *existieren* können. Ein Schätzwert heisst dann „erschöpfend", wenn keine weitergehende Möglichkeit existiert, eine zusätzliche Information über den wahren Wert von θ zu erhalten; diese sind aber äusserst selten, da sie nur unter ganz bestimmten, sehr einschränkenden Bedingungen existieren; $\bar{x}$ und s^2 sind z. B. erschöpfende Schätzwerte von ξ und σ^2 der Normalverteilung (12.1); für nähere Einzelheiten siehe KENDALL [A], Band II, S. 8ff.

Diese hier nur kurz angedeutete Max. Likelihood-Methode führt für die PEARSON III-Verteilung (1) auf folgende *Schätzwerte* $\hat{\alpha}$ und $\hat{\gamma}$:

$$\text{Aus (A) folgt: } L(x_1, \ldots x_n; \alpha, \gamma) = \frac{\alpha^{n\gamma}}{[\Gamma(\gamma)]^n} \left(\prod_{j=1}^{n} x_j^{\gamma-1} \right) e^{-\alpha \sum_{j=1}^{n} x_j}$$

$$\text{aus (B): } (\alpha)\ \frac{\partial \ln L}{\partial \alpha} = \frac{n\gamma}{\alpha} - \sum_1^n x_j = 0$$

$$\frac{\Gamma'(\gamma)}{\Gamma(\gamma)} = \frac{d}{d\gamma}[\ln \Gamma(\gamma)]; \quad (\beta)\ \frac{\partial \ln L}{\partial \gamma} = n \ln \alpha - n \frac{\Gamma'(\gamma)}{\Gamma(\gamma)} + \sum_1^n \ln x_j = 0$$

Die Auflösung von (α) führt auf den gleichen Schätzwert wie bei der Momentenmethode:

$$\hat{\alpha} = \alpha^* = \frac{\gamma}{\bar{x}} \qquad (12)$$

(β) ist nicht explizit nach γ auflösbar. $\hat{\gamma}$ ist die *einzige positive Wurzel der transzendenten Gleichung:*

$$\frac{d}{d\gamma}[\ln \Gamma(\gamma)] = \frac{1}{n}\sum_{1}^{n} \ln x_j + \ln \alpha \tag{13}$$

Ist aber z.B. $\gamma > 10$ – die Grössenordnung von γ kann ja durch den Momentenschätzwert (10) bestimmt werden – so können wir für $\frac{d}{d\gamma} \ln \Gamma(\gamma)$ die Ableitung der in der Analysis häufig gebrauchten *asymptotischen Formel*

$$\ln \Gamma(\gamma) = \ln \sqrt{2\pi} + (\gamma - \tfrac{1}{2}) \ln \gamma - \gamma + \left(\frac{1}{12\gamma} - \frac{1}{360\gamma^3} + \ldots\right)$$

$$\frac{d}{d\gamma} \ln \Gamma(\gamma) = \ln \gamma - \frac{1}{2\gamma} - \frac{1}{12\gamma^2} + \left(\frac{1}{120\gamma^4} + \ldots\right)$$ in (13) einsetzen.

Berücksichtigen wir nur die 2 ersten Glieder, und setzen wir nach (12)

$$\ln \alpha = \ln \gamma - \ln \bar{x},$$

so kann (13) nach γ aufgelöst werden:

$$\hat{\gamma} = \frac{1}{2\left[\ln \bar{x} - \frac{1}{n}\sum_{j=1}^{n} \ln x_j\right]} \tag{13a}$$

Interessant ist, dass hier anstatt der Streuung $s^2(x)$ wie beim Momentenschätzwert eine neue Stichprobengrösse:

$$\frac{1}{n}\sum_{1}^{n} \ln x_j = \ln x_G$$

(x_G = *geometrisches Mittel* der transformierten Grössen x)

in die Formel für $\hat{\gamma}$ eingeht; demnach ist $\hat{\gamma} > 0$, da für $x_i > 0$ immer gilt: $\bar{x} > x_G$, also auch $\ln \bar{x} > \ln x_G$.

135 Durchschnitt μ(y) *der Originalgrundgesamtheit* g(y); *seine Schätzung durch die transformierten Grössen* x = log y.

$$\mu(y) = \int_{1}^{\infty} y\,g(y)\,dy = \int_{0}^{\infty} 10^x f(x)\,dx$$

$$\rightarrow \mu(y) = \left(1 - \frac{1}{\alpha \log e}\right)^{-\gamma} \tag{14}$$

Je nachdem wir die Momenten- oder die Max. Likelhood-Schätzwerte für α und γ in (14) einsetzen, erhalten wir zwei verschiedene („indirekte") Schätzungen $\mu^*(y)$ und $\hat{\mu}(y)$ von $\mu(y)$:

$$\mu^*(y) = \left(1 - \frac{1}{\alpha^* \log e}\right)^{-\gamma^*} \tag{15}$$

wobei α^* und γ^* durch (10) und (11) gegeben sind.

$$\hat{\mu}(y) = \left(1 - \frac{1}{\hat{\alpha} \log e}\right)^{-\hat{\gamma}} \tag{16}$$

$\hat{\alpha}$ und $\hat{\gamma}$ durch (12) und (13) oder (13a) bestimmt.

Um die Streuung von $\mu^*(y)$ bezw. $\hat{\mu}(y)$ schätzen zu können, hätten wir aus den (approximativen) Parameterverteilungen auf die angenäherten Verteilungen von $\mu^*(y)$ oder $\hat{\mu}(y)$ zu schliessen. Die diesbezüglichen, vom Verfasser entwickelten Ableitungen und Resultate seien hier nicht wiedergegeben, da sie recht langwierig sind; ausserdem wird es sich zeigen, dass die Annahme der PEARSON III-Verteilung den Durchschnitt wiederum unterschätzt.

136 Überprüfung der Hypothese 132 am Untersuchungsmaterial

Aus dem uns zur Verfügung stehenden Erfahrungsmaterial wurden wiederum die *Autohaftpflichtschäden des Jahres 1950* herangezogen. Den empirischen, absoluten Häufigkeiten f_j^{1950} wurden die theoretischen, diesmal nach *Ansatz (13.1) berechneten Häufigkeiten* φ_j^{1950} gegenübergestellt; die Güte der Anpassung konnte dann mit Hilfe des χ^2-Testes beurteilt werden (Anhangtabelle 1, Kolonnen (5) und (7)).

Mit $\bar{x} = 2{,}32333$ und $s^2(x) = 0{,}34869$ ergab die Auswertung nach (10) bezw. (11) folgende *Momentenschätzwerte für* α *und* γ:

$$\gamma^* = 15{,}48 \qquad \alpha^* = 6{,}67$$

Zur Ermittlung des *Max. Likelihood-Schätzwertes* $\hat{\gamma}$ konnte die Näherungsformel (13a) angewandt werden, da im vorliegenden Fall $\gamma > 10$ ist; zur Bestimmung von $\ln x_G = \frac{1}{n} \sum_1^l f_j \ln x_j$ mussten zusätzlich die Logarithmen der Klassenmitten x_j gebildet werden; es resultierte für

$$\ln x_G = \ln 10 \cdot \log x_G = \ln 10 \left(\frac{1}{n} \sum_1^l f_j \log x_j\right) = 0{,}81083$$

und: $\ln \bar{x} = \ln 10 \,.\, \log \bar{x} = 0{,}84300$

und demnach nach (13a):

$$\hat{\gamma} = 15{,}54 \qquad \hat{\alpha} = 6{,}67 \text{ nach (12)}$$

Zur Berechnung der φ_j^{1950} wurden die PEARSON'schen Tabellen der unvollständigen Gamma-Funktion [20] verwendet. Für die Argumentwerte u und p der dort tabellierten Verteilungsfunktion:

$$I(u, p) = \frac{\int_0^v e^{-v} v^p dv}{\Gamma(p+1)} \; ; \; v = \sqrt{p+1}\, u$$

mussten in unserem Fall sinngemäss folgende Werte eingesetzt werden:

$$u = \frac{\sqrt{\gamma}}{\bar{x}} . x; \quad p = \gamma - 1$$

Zur einfacheren Interpolation wurde mit einem mittleren $\bar{\gamma} = 15{,}5$ gerechnet; aus den Tabellen wurden die den Klassengrenzen $u_k = \frac{\sqrt{\gamma}}{\bar{x}} x_k$ entsprechenden interpolierten Wahrscheinlichkeitswerte $I(u_k, \gamma - 1)$ entnommen und daraus die Klassenhäufigkeiten $\varphi_j^{1950} = n^{1950} [I(u_{k+1}, \gamma - 1) - I(u_k, \gamma - 1)]$ gewonnen. (Siehe *Anhangtabelle 1, Kolonne (7)* sowie *Fig. 2*).

Die erstrebte Anpassung an die empirischen f_j ist schon bedeutend besser als beim Ansatz der Normalverteilung (12.1), was auch der χ^2-*Test* bestätigt; die Ausrechnung ergab ein

$$\chi^2 = 25{,}91 \text{ mit } N = 17 \text{ } \textit{Freiheitsgraden.}$$

Die Wahrscheinlichkeit $P(\chi^2)$ des theoretisch möglichen Auftretens eines grösseren χ^2 beträgt *rund* 8%, so dass die Abweichungen zwischen theoretischer und beobachteter Verteilung noch als knapp zufällig beurteilt werden können.

Mit $\bar{\gamma} = 15{,}5$ und $\bar{\alpha} = 6{,}67$ ergab sich für den *Durchschnitt* $\bar{\mu}(y)$ gemäss (15) und (16):

$$\bar{\mu}(y)^{1950} = 707$$

direkter Durchschnitt: $\bar{y}^{1950} = 857$

Trotzdem $\bar{\mu}^{1950}(y)$ dem „wahren" Wert schon bedeutend näher kommt als *der* FINNEY-*Wert* m_y^{1950} *in 126*, fällt die Abweichung von $\bar{y}^{1950}$ noch immer ins Gewicht; es ist fraglich, ob diese Diskrepanz

allein auf die Zufallsschwankungen hauptsächlich der grossen Schäden zurückgeführt werden darf.

Um diese Frage abzuklären, wurde es für nötig erachtet, das Material eines andern Jahres dahin zu untersuchen; mit Absicht wurde das *Jahr 1949* gewählt, da dieses einen günstigen Schadenverlauf mit wenig Gross-Schäden aufwies, ganz im Gegensatz zu 1950, das ein extrem schlechtes Jahr war mit einigen sehr bedeutenden Schäden, die $\bar{y}^{1950}$ massgeblich beeinflussten. Wenn die Bestimmung von $\mu(y)$ nach (15) oder (16) befriedigen und die Zufallsschwankungen des „wahren" Durchschnittes besser abschätzen soll als das direkt ermittelte $\bar{y}$, so *müsste* demnach für das ausserordentlich günstige Jahr 1949 ein $\mu^{1949}(y) > \bar{y}^{1949}$ resultieren.

Das statistische Material von 1949 wurde in analoger Weise wie für 1950 bereitgestellt und ausgewertet; die f_j^{1949} und φ_j^{1949} finden sich in Anhangtabelle 1, Kolonnen (8) und (9), ihre bildliche Veranschaulichung in Fig. 2. Es ergaben sich folgende Werte für die Parameter α und γ von (1) und die entsprechenden Schätzwerte für $\mu(y)$:

$$n^{1949} = 737 \quad \hat{\gamma} = 17{,}139 \qquad \gamma^* = 16{,}884$$
$$\hat{\alpha} = 7{,}366 \qquad \alpha^* = 7{,}256$$
$$\hat{\mu}(y)^{1949} = 616{,}6 \qquad \mu^*(y)^{1949} = 629{,}9$$
$$\bar{y}^{1949} = 680{,}2$$

Entgegen unserer Erwartung bleibt nun demnach

$$\bar{y}^{1949} > \mu(y)^{1949}$$

Trotz der stark verbesserten Anpassung müssen wir wiederum wie nach 12 den bedauerlichen Schluss ziehen, dass die *Hypothese der logarithmischen* Pearson *III-Verteilung als Grundgesamtheit der Autohaftpflicht-Schaden-Verteilung für eine gute Schätzung des durchschnittlichen jährlichen Schadenbetrages* M[1] ungeeignet ist, da dieser meistens unterschätzt wird.

14 FOLGERUNGEN AUS DEN ANALYTISCHEN LÖSUNGSVERSUCHEN

Die in diesem Kapitel unternommenen Versuche einer *analytischen Fassung* für die Verteilung der Schadenbeträge in der Autohaftpflichtversicherung brachten wohl eine sukzessiv verbesserte Anpassung.

Besonders der PEARSON III-Ansatz kann bestimmt für gewisse Zwecke befriedigen; so könnte diese Häufigkeitsfunktion und eventuell auch die PARETO-Funktion bei *risikotheoretischen Studien* in allen Haftpflichtversicherungsarten als angenäherte, analytische Risikosummenverteilung dienen, besonders wenn die einzelnen Schadensummen auf eine geeignete mittlere Risikosumme als Einheit bezogen werden, z. B. auf das *geometrische Mittel* der Schadenbeträge, das, wie wir im nächsten Kapitel sehen werden, den Zufallsschwankungen bedeutend weniger ausgesetzt ist als der Schadendurchschnitt. Für solche Untersuchungen ist es empfehlenswert, das Zahlenmaterial von mehreren Jahren zusammen zu verarbeiten, um zuverlässige Werte zu gewinnen.

Wir aber suchen aus dem beschränkten Jahresmaterial einer Gesellschaft aus den häufigsten kleinen und mittleren Schäden auf einen verlässlichen durchschnittlichen Schadenbetrag zu schliessen, der von den seltenen grossen Schäden entscheidend beeinflusst wird, und der dennoch weniger schwanken soll als der gewöhnliche Jahresdurchschnitt. Sowohl die *lognormale* als auch die *logarithmische PEARSON III-Häufigkeitskurven vermögen sich dem Verlauf der Gross-Schäden zu wenig anzupassen, so dass der aus ihnen ermittelte Durchschnitt einen zu niedrigen Schätzwert des wahren Durchschnittes ergibt.* Bis jetzt ist es den verschiedenen Autoren nicht gelungen, ein explizites analytisches Gesetz zu finden, das unser Problem zufriedenstellend löst.

Deshalb verzichten auch wir im nächsten Kapitel auf die analytische Darstellung der Häufigkeitsfunktion der Gesamtheit der Körper- und Sachschadenbeträge, wobei eine gesonderte Behandlung dieser zwei diese Gesamtheit bildenden Komponenten, wie schon erwähnt, wünschenswert gewesen wäre und wahrscheinlich, besonders bei den Sachschäden, gute Resultate ergeben hätte. Leider war ein solches Vorgehen nicht möglich, da in den uns zur Verfügung stehenden Schadenregistern bei den Schadenbeträgen nicht vermerkt war, ob diese die Folgen eines Sach- oder Personenschadens waren.

Wir erachten es deshalb als vorteilhaft, im Folgenden ein direkteres Verfahren zu entwickeln, wobei wir uns auf einige Erkenntnisse und Resultate dieses Kapitels stützen und sie sinngemäss übertragen. Aus diesem Grunde hat sich eine etwas ausführlichere Behandlung der analytischen Methoden trotz des Misserfolges gelohnt.

2 Grundzüge einer neuen Methode zur Bestimmung des durchschnittlichen Schadenbetrages

21 DIE GRUNDLEGENDE IDEE. DAS GEOMETRISCHE MITTEL ALS CHARAKTERISTISCHE GRÖSSE

Den bisherigen Ausführungen entnehmen wir, dass grundsätzlich eine theoretische, hypothetische Grundgesamtheit, aus der unsere beobachteten Schäden Zufallsstichproben bilden, für die Häufigkeiten der Schadenbeträge existiert. Zur Bestimmung des durchschnittlichen Schadenbetrages hat sich aber das Aufsuchen ihrer genauen analytischen Form als schwierig und unvorteilhaft erwiesen; überdies kann man noch beifügen, dass einerseits die Zahlen nur eines Geschäftsjahres nicht sehr zuverlässige Werte für die Kenngrössen dieser mathematisch formulierten Grundgesamtheit ergeben, dass aber anderseits durch die Zusammenfassung des Materials mehrerer Jahre die zeitliche Entwicklungstendenz des die *Lage* der Verteilung charakterisierenden Parameters verwischt wird.

Unsere *neue, grundlegende Hypothese* ist nun die folgende: Beziehen wir die einzelnen Schäden auf irgendeine *charakteristische Jahresgrösse*, von der wir fordern, dass sie

1. die *Höhe aller eingetretenen Schäden des Jahres als Ganzes* wiedergibt und
2. die *zeitliche Entwicklung der Schäden von Jahr zu Jahr* berücksichtigt,

so sollen die *beobachteten Schäden alle als Stichproben aus* der gleichen, *zeitinvarianten Grundgesamtheit* angesehen werden können. Es handelt sich um eine ähnliche Überlegung, wie sie in der kollektiven Risikotheorie angestellt wird. Die Erfahrung zeigte dort, dass die Risikosummenverteilung im Laufe der Zeit praktisch konstant bleibt, wenn als Masseinheit die durchschnittliche Risikosumme gewählt wird. Siehe z. B. Ammeter [21].

Auf der Suche nach der für unser Problem geeigneten, charakteristischen Grösse fiel die Wahl bald auf das *geometrische Mittel* y_G^i *der Schäden des Jahres* i, das ersichtlich den zwei obigen Forderungen entspricht und gegenüber dem Durchschnitt $\bar{y}^i$ grosse Vorteile aufweist; so ist der Einfluss der grossen Schäden auf y_G^i bedeutend kleiner als auf $\bar{y}^i$, womit die Zufallsschwankungen erheblich reduziert und besser abgeschätzt werden können; praktisch lässt sich ausserdem y_G^i und dessen Verteilung relativ leicht ermitteln, wenn wir einmal eine logarithmische Argumentenfolge analog der in 126 erörterten festgelegt haben; die dort vorgenommene Klasseneinteilung werden wir weiterhin benützen; es gilt ja dann für den Durchschnitt $\bar{x}$ der logarithmisch transformierten Schadengrössen $x = \log y$:

$$\bar{x} = \frac{1}{n} \sum_{k=1}^{n} f_k x_k = \frac{1}{n} \sum_{k=1}^{n} f_k \log y_k = \log y_G \qquad (1)$$

Somit wird uns schon rein intuitiv y_G^i als die die Lage der Verteilung kennzeichnende Grösse nahegelegt.

Ausserdem kann noch darauf hingewiesen werden, dass das geometrische Mittel dann als charakteristische Grösse einer Verteilung angemessen ist, wenn wir es mit Elementen zu tun haben, deren Änderungen direkt proportional zu den Elementen selbst sind. Das arithmetische Mittel eignet sich bekanntlich dann besonders gut zur Charakterisierung einer Verteilung, wenn die Abweichungen von ihm näherungsweise das Gauss'sche Gesetz befolgen, was aber im vorliegenden Falle ja gerade nicht zutrifft, wie Kapitel 1 ergeben hat.

Falls sich die Hypothese der Zeitinvarianz der auf das geometrische Mittel y_G^i bezogenen Elemente der Grundgesamtheit an unserem Erfahrungsmaterial bestätigt (siehe 22), so bestimmen sich *der durchschnittliche Schadenbetrag* M^i *und dessen Streuung* $\sigma(M^i)$ aus den folgenden Beziehungen:

$$M^i = Q^i \cdot y_G^i \text{ oder: } \log M^i = \log Q^i + \log y_G^i \qquad (2)$$

wenn wir durch

$$Q^i = \frac{y^i}{y_G^i} \qquad (3)$$

den *sogenannten „Mittelwertquotienten" definieren.*

Q^i *bezw.* $\log Q^i$ ist dann nach unserer fundamentalen Hypothese eine *zeitinvariante Grösse* Q *bezw.* $\log Q$, die nun unbedenklich aus

einem mehrere Jahre umfassenden Beobachtungsmaterial durch Stichprobenbildung ein für alle Mal ermittelt werden kann.

Aus plausiblen Annahmen über die Häufigkeitsfunktionen von $\log y_G^i$ (23) und $\log Q$ (24) kann dann auf $\log M^i = \log Q + \log y_G^i$, bezw. M^i und *dessen Streuung* $\sigma(M^i)$ geschlossen werden (25).

22 HOMOGENITÄTSTESTE ZUR BESTÄTIGUNG DER HYPOTHESE DER ZEITINVARIANZ DES MITTELWERTQUOTIENTEN Q^i

221 Theoretische Grundlage:

Die zu prüfende Hypothese ist die folgende: Wird das *geometrische Mittel* y_G^i aus den Schäden des Jahres i *als Masseinheit* der Zufallsvariablen y_k (= Schadenhöhe) gewählt:

$$Y_k = \frac{y_k}{y_G^i} \quad , \quad (k = 1, 2, \ldots, r) \tag{1}$$

so bilden die so gewonnenen *beobachteten, absoluten Häufigkeiten* f_k^i ($i = 1, 2, \ldots, s$; $k = 1, 2, \ldots, r$) s empirische Häufigkeitsverteilungen der Jahre $i = 1, 2, \ldots, s$; diese können als s *unabhängige Stichproben* aus der *gleichen* hypothetischen *Grundgesamtheit* angesehen werden, d.h. diese *Daten* f_k^i sind *dann innerhalb der verschiedenen Jahre* i *homogen.*

Dies ist der Hypothese äquivalent, dass r *Konstanten* $p_1, p_2, \ldots p_r$ existieren mit $\sum_{k=1}^{r} p_k = 1$, so dass die Wahrscheinlichkeit, dass eine beobachtete Grösse in die k-te Klasse Y_k fällt, in jeder der s Stichproben p_k beträgt. Diese r Wahrscheinlichkeiten bilden dann gewissermassen die zeitinvarianten Parameter der Häufigkeitsverteilung der unbekannten Grundgesamtheit.

Wir können unser gesamtes Beobachtungsmaterial, das zur Bestätigung dieser Hypothese dienen soll, in das folgende *Schema* einordnen, das das Bild einer sogenannten Mehrfeldertafel aufweist:

Klassen k	*Jahre i* 1	2 . . . . . i . . . . . s	*Total*	
1	f_1^1	f_1^2 . . . . . f_1^s	n_1	
2	f_2^1	f_2^2 . . . . . f_2^s	n_2	
.	.	. .	.	
.	.	. .	.	
.	.	. .	.	
k	f_k^1	f_k^2 . . . f_k^i . . . f_k^s	n_k	$p_k = \frac{n_k}{n}$
.	.	. . .	.	
.	.	. . .	.	
.	.	. . .	.	
r	f_r^1	f_r^2 . . . . . f_r^s	n_r	
	n^1	n^2 . . . n^i . . . n^s	n	

Die *Prüfung der aufgestellten Hypothese* kann *mit Hilfe des* χ^2*-Testes* erfolgen. Nach Cramer [B] S. 443–450 unterliegt die Grösse:

$$\chi^2 = \sum_{i,k} \frac{(f_k^i - n^i p_k)^2}{n^i p_k} = n \sum_{i,k} \frac{\left(f_k^i - \frac{n^i n_k}{n}\right)^2}{n^i n_k}$$

$$= n\left(\sum_{i,k} \frac{(f_k^i)^2}{n^i n_k} - 1\right) \qquad (2)$$

einer χ^2*-Verteilung mit* $(r - 1) \cdot (s - 1)$ *Freiheitsgraden.*

222 Praktische Auswertung: (Anhangtabelle 2)

Vorerst wurden die *empirischen Häufigkeitsverteilungen* der Autohaftpflicht-Schadenbeträge der *Jahre 1947–1951* und zu Vergleichszwecken diejenige des weit zurückliegenden Jahres *1938* auf analoge Weise den Schadenregistern dieser Jahre entnommen, wie wir dies in 126 ausführlich für 1950 erläutert haben; es wurden also vorläufig die gleichen Klassengrenzen und -mitten wie die in Anhangtabelle 1 wiedergegebenen zugrunde gelegt. Aus der Beziehung:

$$\log y_G^i = \frac{1}{n^i} \sum_k f_k^i \log y_k$$

wurden dann die *geometrischen Mittel* y_G^i der Jahre 1938, 1947–1951 gewonnen. Es ergaben sich folgende Werte:

Tabelle 2a:

Jahr i	n^i	$\log y_G^i$	y_G^i
1938	404	2,15965	144,4
1947	559	2,31664	207,3
1948	723	2,33103	214,3
1949	737	2,32678	212,2
1950	968	2,32333	210,5
1951	969	2,36154	229,9

Nun konnten die mit y_G^i *normierten Klassengrenzen* Y_k gemäss (221.1) gebildet werden. Ausgehend von $y_G^{1950} = 210{,}5$ erhielt man die in Anhangtabelle 2 ersichtlichen

$$Y_k = \frac{y_k}{y_G^{1950}}$$

wobei y_k die ursprünglichen Klassengrenzen sind (Anhangtabelle 1, Kolonne (3)); die entsprechenden Klassenhäufigkeiten f_k^{1950} konnten dann direkt aus Tabelle 1, Kolonne (5) übertragen werden.

Um überhaupt einen Anhaltspunkt über die Richtigkeit unserer Hypothese zu haben, wurde zunächst der *Homogenitätstest (2)* auf die Daten *der zwei weit auseinanderliegenden Jahre 1938 und 1950* angewandt (s = 2). Zur Ermittlung der f_k^{1950} mussten die *modifizierten Klassengrenzen:*

$$y_k' = y_G^{1938} \cdot Y_k$$

gebildet werden; eigentlich hätten wir dann die Schäden 1938 unter Beachtung der neuen Klassengrenzen y_k' neu zu gruppieren. In unserem Falle ergab sich insofern eine Vereinfachung, als dass sich mit $y_G^{1938} = 144{,}4$ fast die gleichen, gegenüber den ursprünglichen y_k nur um eine Klasse nach unten verschobenen Klassengrenzen y_k' ergaben. Dies erklärt sich damit, dass $\frac{y_G^{1950}}{y_G^{1938}} \sim 1{,}46$, also praktisch gleich dem konstanten Verhältnis q zweier aufeinanderfolgender Argumentwerte y ist (siehe 126). Die f_k^{1938} konnten somit auch aus der ursprünglichen Statistik sinngemäss übernommen werden.

χ^2-Test für s = 2 *(1938, 1950) gemäss (221.2)*

Die Anwendung des χ^2-Testes erfordert eine Besetzung von mindestens 6 Elementen in einer Klasse. Dies wurde durch Zusammen-

fassung der ersten 4 und der letzten 8 Gruppen (siehe Anhangtabelle 2, Daten 1938 und 1950) erreicht, somit ergaben sich r = 17 *Klassen.*

Die Rechung lieferte ein

$$\chi^2 = 11{,}56 \text{ mit } N = (r - 1) \cdot (s - 1) = 16 \textit{ Freiheitsgrade}$$

Wahrscheinlichkeit einer grösseren Abweichung:

$$P(\chi^2) = 0{,}77$$

Wir haben also eine überraschend gute Bestätigung der aufgestellten Hypothese, da die Abweichungen zwischen den empirischen Verteilungen 1938 und 1950 nur zufällig sind.

Um ganz sicher zu gehen, wurde nun die Homogenität des *gesamten Beobachtungsmaterials der Jahre 1938, 1947–1951 mit Hilfe des Testes (221.2) geprüft.* Da die geometrischen Mittel der Jahre 1947–51 nicht stark voneinander abweichen, wurden die ursprünglichen, zur Bestimmung von y_G^i benutzten Häufigkeiten f_k^i eingesetzt, wobei wir uns des dabei gemachten Fehlers bewusst sind, diesen aber der Einfachheit halber in Kauf nehmen (Anhangtabelle 2). Wir haben s = 6 *unabhängige Stichproben:* durch Zusammenfassung der schwach besetzten Randklassen wie oben r = 17 *Gruppen.* Die umfangreiche rechnerische Auswertung (221.2) ergab ein

$$\chi^2 = 79{,}72 \text{ mit } N = (r - 1) = 80 \textit{ Freiheitsgraden}$$

Die Grösse $z = \sqrt{2\chi^2} - \sqrt{2N - 1} = 0{,}0174$ ist für $N > 30$ bekanntlich in guter Näherung standardisiert normalverteilt; aus einer Tafel der Gauss'schen Fehlerfunktion resultierte eine Wahrscheinlichkeit:

$$P(\chi^2) = 1 - \Phi(z) = 0{,}489$$

Damit sehen wir in schöner Weise die Vermutung der *zeitlichen Invarianz der mit* y_G^i *normierten Häufigkeitsverteilung der Grundgesamtheit* bestätigt.

Daraus folgt aber sofort die *zeitliche Konstanz* unseres „*Mittelwertquotienten*"

$$Q^i = \frac{\bar{y}^i}{y_G^i} = Q \qquad (3)$$

23 DIE HÄUFIGKEITSVERTEILUNG DES GEOMETRISCHEN MITTELS y_G^i BEZW. $\log y_G$

In 13 haben wir gesehen, dass die transformierten Schadenbeträge $x = \log y$ angenähert PEARSON III-verteilt sind gemäss (13.1); zur Ermittlung des durchschnittlichen Schadenbetrages war die Annahme dieser Häufigkeitsfunktion als Grundgesamtheit wegen der ungenügenden Berücksichtigung der grossen Schäden aber zu ungenau. Zur *theoretischen Herleitung der Häufigkeitsverteilung von* $\log y_G^i$ postulieren wir daher, $x = \log y$ sei PEARSON III-verteilt. Dabei wissen wir, dass dies eine recht gute Näherung darstellt; ausserdem sind ja erfahrungsgemäss (Tabelle 2a) die Unterschiede von y_G^i innerhalb der verschiedenen Jahre unbedeutend im Vergleich zu den Jahresschwankungen von $\bar{y}^i$. Dies deutet aber darauf hin, dass y_G^i viel weniger von den grossen Schäden beeinflusst wird als $\bar{y}^i$. Somit sind wir nun leicht in der Lage, die *Verteilung von* $\bar{x}^i = \log y_G^i$ aus Stichproben vom Umfang n^i wie folgt zu gewinnen:

Nach (13.3) ist
$$\varphi(t) = \left(1 - \frac{it}{\alpha}\right)^{-\gamma} \quad (i = \sqrt{-1})$$

die der PEARSON III-Verteilung (13.1):

$$f(x) = \frac{a^\gamma}{\Gamma(\gamma)} e^{-\alpha x} x^{\gamma-1}$$

zugeordnete charakteristische Funktion. Nach dem Produktsatz über charakteristische Funktionen ist dann die charakt. Funktion der Summe der n zufälligen, stochastisch unabhängigen Variablen

$$\sum_{j=1}^{n^i} x_j = n^i \bar{x}^i: \quad \left(1 - \frac{it}{\alpha}\right)^{-n^i\gamma}$$

Die charakt. Funktion bezüglich $\bar{x}^i$ folgt aus dem Transformationssatz:

$$\left(1 - \frac{it}{n^i\alpha}\right)^{-n^i\gamma}$$

Daraus schliessen wir nach dem Inversionssatz sofort auf die *Stichproben-Häufigkeitsverteilung von* $\bar{x}^i = \log y_G^i$:

$$f(\bar{x}^i) = f(\log y_G^i) = \frac{(n^i\alpha)^{n^i\gamma}}{\Gamma(n^i\gamma)} e^{-\alpha n^i \bar{x}^i} (\bar{x}^i)^{n^i\gamma-1} \qquad (1)$$

Damit ergeben sich folgende *Kenngrössen bezüglich (1):*

aus (13.6) → für den *Mittelwert* $\mu(\bar{x}^i) = \frac{\gamma}{\alpha} = \mu(x)$ (2)

aus (13.7) → für die *Streuung* $\mu_2(\bar{x}^i) = \sigma^2(\bar{x}^i) = \frac{\gamma}{\alpha^2 n^i} = \frac{\sigma^2(x)}{n^i}$ (3)

aus (13.8) → für die *Schiefe* $\gamma_1(\bar{x}^i) = \frac{2}{\sqrt{n^i \gamma}} = \frac{\gamma_1(x)}{\sqrt{n^i}}$ (4)

aus (13.9) → für den *Exzess* $\gamma_2(\bar{x}^i) = \frac{6}{n^i \gamma} = \frac{\gamma_2(x)}{n^i}$ (5)

Die *Stichprobengrösse* $\bar{x}^i = \log y_G^i = \frac{1}{n^i} \sum_{j=1}^{n^i} x_j$ ist bei der Annahme der PEARSON III-Verteilung als Grundgesamtheit der $x = \log y$ ebenfalls PEARSON III-verteilt; der Erwartungswert $E(\bar{x}^i)$ entspricht dem Durchschnitt $\mu(x)$ der Grundgesamtheit und wird durch $\bar{x}^i$ selbst geschätzt, die mittlere quadratische Abweichung $\sigma(\bar{x}^i)$ beträgt $\frac{\sigma(x)}{\sqrt{n^i}}$ wobei das *unbekannte* $\sigma(x)$ der Grössen $x = \log y$ durch

$$s(x) = \sqrt{\frac{1}{n^i - 1} \sum_{j=1}^{n^i} (x_j - \bar{x}^i)^2} \quad \text{geschätzt wird.}$$

Aus (4) und (5) erkennen wir, dass die *Verteilung von* $\log y_G^i$ für grössere n^i und γ ziemlich rasch gegen die Normalverteilung strebt.

Bei unseren Stichprobenumfängen n^i können wir deshalb ohne weiteres die *Stichproben-Häufigkeitsverteilung von* $\bar{x}^i = \log y_G^i$ durch *die Normalverteilung* $\left(\mu(x), \frac{\sigma^2(x)}{\sqrt{n^i}}\right)$ *approximieren.*

Den Ausführungen in 12 entnehmen wir dann sofort, dass y_G^i *selbst lognormalverteilt* ist; Schiefe und Exzess dieser Häufigkeitsverteilung werden sehr klein wegen (12.8); das dort definierte

$$\tau = e^{\sigma^2(\bar{x}^i)} = e^{\frac{\sigma^2(x)}{n^i}}$$

liegt bei unseren Stichprobenumfängen n^i sehr nahe bei 1. Bezeichne ich das geometrische Mittel der Grundgesamtheit mit η_G, so fallen wegen $\tau \sim 1$ dichtester Wert $= \eta_G \tau^{-1}$, Zentralwert $\eta_Z =$ geometrisches Mittel η_G und Durchschnitt $= \eta_G \tau^{\frac{1}{2}}$ der lognormalen Häufigkeitsverteilung von y_G^i fast zusammen, so dass wir in guter Näherung setzen können:

$$E(y_G^i) = \mu(y_G^i) = y_G^i = \text{antilog } \bar{x}^i \qquad (6)$$

und für das „*Abweichungsverhältnis*“ $\sigma_L(y_G^i)$, das $\sigma(\bar{x}^i)$ entspricht:

$$\sigma_L(y_G^i) = \text{antilog}\frac{\delta(x)}{\sqrt{n^i}} \tag{7}$$

24 BESTIMMUNG DES MITTELWERTQUOTIENTEN Q UND DESSEN STREUUNG MITTELS STICHPROBEN

Es ist auf *theoretischem Wege* nicht möglich, für Q eine angenäherte Häufigkeitsverteilung in analoger Weise herzuleiten, wie wir dies im vorangehenden Abschnitt 23 für y_G^i tun konnten. Hätten wir nämlich eine befriedigende analytische Verteilung für den Durchschnitt im Zähler des Ausdruckes (222.3) gefunden, so wäre doch unser Problem schon im 1. Kapitel bei den analytischen Lösungsversuchen gelöst gewesen.

Wir müssen deshalb *plausible Annahmen* treffen; weiter sind wir genötigt, aus dem gesamten uns zur Verfügung stehenden Beobachtungsmaterial *Stichproben* zu bilden, um *vor allem die Streuung von* Q *zu ermitteln;* zur Bestimmung von Q allein wäre der Durchschnitt der Q^i aus den 6 Jahren 1938, 1947–1951 ausreichend. Genau so wichtig ist aber für unser Problem *die Schätzung der Zufallsschwankungen des zeitinvarianten* Q. Deshalb muss die Streuung von Q auf empirischem Wege gewonnen werden.

Die aus den m *Stichproben* mit den *Umfängen* m_j *bestimmten Mittelwertquotienten* bezeichnen wir mit

$$q_j^i = \frac{\bar{y}_j^i}{y_G^i} \; (j = 1, 2, \ldots, m) \tag{1}$$

wobei: $\bar{y}_j^i = \frac{1}{m_j} \sum_{l=1}^{m_j} y_l^i$ (i = Jahresindex) (j,l = Summationsindices)

und y_G^i das übliche *geometrische Mittel des Jahres* i bedeutet.

Für die *Häufigkeitsverteilung der* m *Stichprobengrössen* q_j^i werden nun folgende *zwei verschiedene Annahmen* gemacht:

1. Die zufällige Variable $q_j^i (j = 1, 2, \ldots, m)$ ist für $m \to \infty$ eine *normalverteilte* kontinuierliche Zufallsvariable q:

$$E(q) = \bar{q} = Q = \frac{1}{m} \sum_{j=1}^{m} q_j^i \tag{2}$$

$$\sigma^2(q) = s^2(q_j^i) = \frac{1}{m-1} \sum_{j=1}^{m} (q_j^i - \bar{q})^2 \qquad \text{und daraus}$$

$$\sigma(Q) = \frac{\sigma(q)}{\sqrt{m}} = \frac{s(q_j^i)}{\sqrt{m}} \tag{3}$$

2. *Die logarithmisch transformierte zufällige Variable*

$$z_j^i = \log q_j^i = \log \bar{y}_j^i - \log y_G^i$$

ist eine *asymptotisch normalverteilte Zufallsgrösse* z. Dies ist die plausiblere Hypothese als die erste; von $\log y_G^i$ wissen wir, dass sie asymptotisch normalverteilt ist, und mangels einer theoretisch besser begründbaren Annahme postulieren wir, dass auch $\log \bar{y}_j^i$ einer Normalverteilung unterliegt. Als Differenz zweier normalverteilter Veränderlichen ist demnach auch $z_j^i = \log q_j^i$ beim Grenzübergang $m \to \infty$ normalverteilt.

q *selbst* ist dann *lognormalverteilt.*

Für den *Durchschnitt* $\mu(q) = Q$ und die *Streuung* $\mu_2(q) = \sigma^2(q)$ der lognormalen Grundgesamtheit q können infolgedessen die *wirksamsten Schätzwerte nach* Finney *[11] gemäss (125.17) und (125.18)* eingesetzt werden; ist:

$$\bar{z} = \frac{1}{m} \sum_{j=1}^{m} \ln q_j^i = \ln 10 \, \frac{1}{m} \sum_{j=1}^{m} \log q_j^i \tag{A}$$

$$\begin{aligned} s^2(z) &= \frac{1}{m-1} \sum_{j=1}^{m} (\ln q_j^i - \bar{z})^2 \\ &= \ln^2 10 \left\{ \frac{1}{m-1} \left[\sum_{j=1}^{m} (\log q_j^i)^2 - \frac{1}{m} \left(\sum_{j=1}^{m} \log q_j^i \right)^2 \right] \right\} \end{aligned} \tag{B}$$

so wird: $\mu(q) = Q = e^{\bar{z} + \frac{1}{2} s^2} \left\{ 1 - \frac{s^2(s^2+2)}{4m} + 0\left(\frac{1}{m^2}\right) \right\}; \; s^2 = s^2(z)$ (4)

$$\mu_2(q) = \sigma^2(q) = e^{2\bar{z} + s^2} \left\{ e^{s^2} \left[1 - \frac{2s^2(2s^2+1)}{m} + 0\left(\frac{1}{m^2}\right) \right] - \left[1 - \frac{s^2(s^2+2)}{m} + 0\left(\frac{1}{m^2}\right) \right] \right\}$$

und daraus für kleinere s^2 und grössere m:

$$\mu_2(Q) = \sigma^2(Q) = \frac{\sigma^2(q)}{m} \sim \frac{e^{2\bar{z}+s^2}(e^{s^2}-1)}{m} \qquad (5)$$

Q *ist unter der getroffenen Annahme normalverteilt mit einem Durchschnitt (4) und einer Streuung (5).*

In Analogie zum vorderen Abschnitt 23 und im *Hinblick auf* die *Zusammensetzung von* M^1 *aus* y_G^i *und* Q ist es wünschenswert nicht für Q selbst, sondern für log Q *eine angenäherte Normalverteilung* zu bekommen. Wir können eine solche folgendermassen gewinnen:

Aus (4) folgt näherungsweise:

$$\ln Q = \bar{z} + \tfrac{1}{2} s^2$$

$\bar{z}$ ist bekanntlich als Summe von normalverteilten Zufallsvariablen z_i^i wieder normalverteilt mit:
$\mu(\bar{z}) = E(\bar{z}) = \zeta$, wo ζ der Durchschnitt der Grundgesamtheit z ist $(m \to \infty)$ und *durch* $\bar{z}$ *gemäss (A) geschätzt* wird.
$\mu_2(\bar{z}) = \sigma^2(\bar{z}) = \frac{\sigma^2(z)}{m}$ wo $\sigma^2(z)$ die durch $s^2(z)$ nach (B) geschätzte Streuung der Grundgesamtheit z ist.
$t = \frac{1}{2} s^2$ ist bekanntlich exakt χ^2-verteilt oder *asymptotisch normalverteilt* mit:

$$\mu(t) = E(t) = \tfrac{1}{2} \sigma^2(z)$$

und

$$\mu_2(t) = \sigma^2(t) = \frac{\sigma^4(z)}{2(m-1)}$$

Ausserdem wissen wir, dass $\bar{z}$ *und* t *stochastisch unabhängig sind.* Es folgt, dass

$\ln Q = \bar{z} + \frac{1}{2} s^2$ *asymptotisch normalverteilt ist mit:*

$$E(\ln Q) = E(\bar{z}) + E[\tfrac{1}{2} s^2] = \zeta + \tfrac{1}{2} o^2(z)$$

$$\mu_2(\ln Q) = \mu_2(\bar{z}) + \mu_2(t) = \frac{\sigma^2(z)}{m} + \frac{\sigma^4(z)}{2(m-1)}$$

und endlich, da: $\log Q = \ln Q \log e$ folgt:

log Q *asymptotisch normalverteilt* mit:

$$E(\log Q) = [\zeta + \tfrac{1}{2} \sigma^2(z)] \log e \qquad (6)$$

$$\sigma^2(\log Q) = \left[\frac{\sigma^2(z)}{m} + \frac{\sigma^4(z)}{2(m-1)}\right] \log^2 e \qquad (7)$$

wobei die Parameter ζ und $\sigma^2(z)$ aus (A) und (B) durch $\bar{z}$ bezw. $s^2(z)$ geschätzt werden.

Q *selbst* ist dann asymptotisch *lognormalverteilt*. Ist $\sigma^2(\log Q)$ relativ klein, so können sinngemäss analoge Folgerungen wie am Schlusse von 23 gezogen werden, d.h. diese Lognormalverteilung ist fast mit einer Gauss'schen Normalverteilung identisch.

25 DIE SCHÄTZUNG DES DURCHSCHNITTLICHEN SCHADENBETRAGES M^1 UND DESSEN STREUUNG

Wir sind nun ohne weiteres in der Lage, die *Häufigkeitsverteilungen von* $\log M^1$ *und von* M^1 anzugeben. Nach (21.3) ist:

$$\log M^1 = \log y_G^i + \log Q$$

og y_G^i ist nach 23 in guter Näherung *normalverteilt* mit:

$$E(\log y_G^i) = \log y_G^i = \frac{1}{n^i} \sum_{l=1}^{n^i} \log y_l^i$$

$$\sigma^2(\log y_G^i) = \frac{\frac{1}{n^i - 1} \sum_{l=1}^{n^i} (\log y_l^i - \log y_G^i)^2}{n^i}$$

Das gemäss 22 von i unabhängige, zeitinvariante $\log Q$ ist nach 24 ebenfalls *asymptotisch normalverteilt* mit:

$E(\log Q) = \log Q = [\bar{z} + \frac{1}{2} s^2(z)] \log e$ nach (24.6) und (A

$\sigma^2(\log Q) = \left[\frac{s^2(z)}{m} + \frac{s^4(z)}{2(m-1)}\right] \log^2 e$ nach (24.7) und (B)

Folglich ist auch $\log M^1$ *angenähert normalverteilt* mit:

$$E(\log M^1) = \log M^1 = E(\log y_G^i) + E(\log Q) \quad (1)$$

$$\sigma^2(\log M^1) = \sigma^2(\log y_G^i) + \sigma^2(\log Q) \quad (2)$$

wenn wir noch $\log y_G^i$ und $\log Q$ als *stochastisch unabhängige Zufallsgrössen* voraussetzen. Diese Annahme dürfte im vorliegenden Fall statthaft sein.

$M^i = y_G^i Q$ selbst ist dann näherungsweise *lognormalverteilt*. Für die Abweichung der Lognormalverteilung von der Normalverteilung ist Schiefe und Exzess nach (12.8), indirekt also die dort definierte Grösse $\tau = e^{\sigma^2(\log M^i)}$ massgebend. Wir werden im nächsten Kapitel sehen, dass die Streuung $\sigma^2(\log M^i)$ ungefähr von der Grössenordnung 0,0015 ist, so dass $\tau \sim 1$, wir also praktisch auch für M^i eine Normalverteilung haben. Wichtig ist dabei die Feststellung, dass sich demnach das geometrische Mittel von $M^i = \eta_G(M^i)$ und der Durchschnitt $\mu(M^i) = \eta_G(M^i)\,\tau^{\frac{1}{2}}$ minim voneinander unterscheiden; wir können also für den *Durchschnitt* und das *Abweichungsverhältnis* des lognormalverteilten *durchschnittlichen Schadenbetrages* M^i *setzen:*

$$E(M^i) = M^i = \text{antilog}\, E(\log M^i) \tag{3}$$

$$\sigma_L(M^i) = \text{antilog}\, \sigma(\log M^i) \tag{4}$$

Eine $k\,\sigma$-fache *Abweichung von* $\log M^i$ mit der entsprechenden Wahrscheinlichkeit $P(k\,\sigma)$, dass

$$\log M^i - k\,\sigma(\log M^i) < \log M^i < \log M^i + k\,\sigma(\log M^i) \tag{5}$$

geht mit der *gleichen Wahrscheinlichkeit* $P(k\,\sigma)$ über in:

$$\sigma_L{}^{-k}(M^i) \,.\, M^i < M^i < \sigma_L{}^{k}(M^i) \,.\, M^i \tag{6}$$

Im letzten Kapitel werden wir nun noch die Zweckmässigkeit der entwickelten Methode an unserem Zahlenmaterial belegen. Besonderes Gewicht wird auf die Bestimmung der grundlegenden Grösse Q gelegt; die verschiedenen Ansätze für Q und log Q in 24 werden verglichen und die dort gemachten Annahmen überprüft. Wenn wir einmal einen zuverlässigen, für das zu untersuchende Portefeuille charakteristischen Mittelwertquotienten Q ermittelt haben, ergeben sich durchschnittlicher Schadenbetrag M^i des Jahres i und weitere Anwendungsmöglichkeiten zwangsläufig und mühelos.

3 Praktische Anwendungen

31 DIE NUMERISCHE ERMITTLUNG DES MITTELWERTQUOTIENTEN Q UND DESSEN STREUUNG

311 Bildung der Stichproben aus dem gesamten Beobachtungsmaterial: Autohaftpflicht-Schäden „Berner Allgemeine", 1938, 1947–1951 (Anhangtabelle 3)

Wir haben schon in 126 bei der praktischen Bestimmung der empirischen Häufigkeitsverteilung der Schadenbeträge des Jahres 1950 aus den Daten des Schadenregisters erwähnt, dass beim Gruppieren der einzelnen Beträge *Untergruppen von je 200 registrierten Schäden* gebildet wurden; dieses Vorgehen war sowohl zu Kontrollzwecken als auch für eventuelle Stichprobenuntersuchungen gedacht, und es wurde in der Folge für die weiteren Jahre beibehalten. Die so gewonnenen Untergruppen mit den entsprechenden *empirischen Häufigkeiten* $f_k^{ij}(k = 1, 2, \ldots, r)$ werden nun als weitere unabhängige Zufallsstichproben aus der hypothetischen Grundgesamtheit angesehen. Eigentlich sind es nicht Stichproben im strengsten Sinne, doch dürfen wir sie mangels Schaden- oder Lochkarten als solche betrachten; die Eintragung der Schäden in das Register erfolgt nämlich in der Reihenfolge der Anmeldung, also zufällig, da in diesem Zeitpunkt der genaue Schadenbetrag noch unbekannt ist. Viele gestellte Haftpflichtansprüche werden abgelehnt oder bleiben später ohne Folgen; da diese ebenfalls registrierten „0-Schäden" bei der Bildung der 200er-Gruppen miteinbezogen wurden, ergaben sich auch *verschiedene Stichprobenumfänge* $m_j = \sum_{k=1}^{r} f_k^{ij}$, die aber nicht allzu sehr schwanken, da der Prozentsatz dieser „0-Schäden" immer ca. 25% beträgt. Wir erhielten einen *durchschnittlichen Stichprobenumfang* $\overline{m}_j = 151$ (Tabelle 3, Kolonne (1)).

Aus dem gesamten Beobachtungsmaterial der 6 Jahre konnten $m = 25$ *ganze Stichproben* gewonnen werden; die Restgruppen in den einzelnen Jahren, die nicht 200 registrierte Schäden ergaben, wurden nicht berücksichtigt. Demnach haben wir *total 3775 Schäden.* Aus den 25 Stichproben ergaben sich die in Tabelle 3, Kolonne (2) wiedergegebenen $\bar{y}_j^i = \frac{1}{m_j} \sum_{k=1}^{r} f_k^{ij} y_k$, wobei die y_k die üblichen Klassenmitten gemäss Tabelle 1, Kolonne (4) sind; für $y_k > 10'000$ Fr. wurden wie immer die exakten Schadenbeträge eingesetzt. y_G^i und $\log y_G^i$ (Kolonnen (4) und (5)) wurden auf bekannte Weise den Jahresbeständen entnommen (Tabelle 2a). Zweckmässigerweise wurde zur *Bestimmung der* q_j^i und $\log q_j^i$ (Kolonnen (6) und (7)) gemäss (241). für die Jahre 1947–1950 ein gemeinsames, durchschnittliches $\bar{y}_G^i$ bezw. $\log \bar{y}_G^i$ gewählt, da sich die geometrischen Mittel dieser Jahre nicht sehr voneinander unterscheiden. Die Logarithmen von $\bar{y}_j^i$ und y_G^i sind zu Kontrollzwecken in der Tabelle 3 aufgeführt.

312 Normalitätsteste für die in 24 getroffenen Hypothesen

In 24 wurde zuerst angenommen, dass die *Stichproben-Mittelwertquotienten* q_j^i *angenähert normalverteilt* sind gemäss (24.2) und (24.3). Um diese Hypothese hauptsächlich im Hinblick auf die Streuung von Q zu überprüfen, wurde ein Normalitätstest angesetzt. Einen solchen, für unsere Zwecke besonders geeigneten *Test* finden wir in Kendall [A], Bd. II, pag. 105, der für *nicht gruppierte Stichprobendaten und auch für mässig kleine Stichprobenumfänge* m *anwendbar* ist.

Demgemäss ist die Häufigkeitsverteilung der Stichprobengrösse:

$$t = \frac{k_3}{k_2^{3/2}}; \quad k_2 = \frac{m}{m-1} m_2(q); \quad m_2(q) = \frac{1}{m} \sum_{j=1}^{m} (q_j^i - \bar{q})^2$$

$$k_3 = \frac{m^2}{(m-1)(m-2)} m_3(q); \quad m_3(q) = \frac{1}{m} \sum_{j=1}^{m} (q_j^i - \bar{q})^3$$

($m_2(q)$ und $m_3(q) = 2.$ und 3. empirische Momente um $m_1' = \bar{q}$) bei der Annahme einer normalverteilten Grundgesamtheit q vom Durchschnitt und der Streuung derselben unabhängig, wiederum angenähert normal verteilt und nur vom Stichprobenumfang m abhängig; die theoretischen Kenngrössen der Verteilung von t lauten:

$$E(t) = \mu(t) = \mu_3(t) = 0$$

$$\mu_2(t) = \sigma^2(t) = \frac{m(m+1)}{(m-2)(m+1)(m+3)}$$

Bilden wir die standardisierte Abweichung $z_0 = \frac{t}{\sigma(t)}$ des berechneten t vom theoretischen Erwartungswert $E(t) = 0$, so kann die Hypothese der normalverteilten Grundgesamtheit q durch eine Wahrscheinlichkeitsaussage beurteilt werden.

Die Auswertung mit Hilfe unserer $m = 25$ *Stichprobengrössen* q_j^i ergab folgende Werte:

(Tabelle 3, Kolonne (6) unten)

$$\left.\begin{array}{l} m_1' = \bar{q} = Q = 3{,}98067 \\ m_2 = m_2' - m_1'^2 = 2{,}68949 \\ m_3 = m_3' - 3m_1'm_2 - m_1'^3 = 7{,}47036 \end{array}\right\} \rightarrow \begin{array}{l} t = 1{,}80475 \\ \sigma^2(t) = 0{,}215 \\ \sigma(t) = 0{,}4637 \end{array}$$

folglich: $z_0 = 3{,}89$ Wahrscheinlichkeit einer grösseren Abweichung: $P(z > z_0) < 0{,}001$

Gemäss diesem Test muss somit die erste Hypothese der normalverteilten Grundgesamtheit q verworfen werden.

Nach der *zweiten Hypothese* sollen die *transformierten Grössen* $z_j^i = \log q_j^i$ *asymptotisch normalverteilt* liegen. Denselben Test auf diese 25 Stichprobengrössen z_j^i angewandt, ergab folgendes Resultat:

$$\left.\begin{array}{l} m_1'(z) = 0{,}57013 \text{ (Tab. 3, Kol. (7) unten)} \\ m_2(z) = 0{,}02381 \\ m_3(z) = 0{,}00287 \end{array}\right\} \rightarrow \begin{array}{l} t = 0{,}83193 \\ \sigma(t) = 0{,}4637 \\ z_0 = 1{,}794 \\ P(z > z_0) = 0{,}073 \end{array}$$

Diese zweite Annahme erweist sich somit als bedeutend besser und plausibler als die erste, trotzdem das Resultat des Testes nicht gerade gut ist. Immerhin dürfen wir sie mangels einer besseren als nahezu gesichert gelten lassen.

313 Gegenüberstellung der nach den verschiedenen Ansätzen in 24 berechneten Mittelwertquotienten

In der folgenden *Tabelle 3a* sind die nach den Ansätzen (24.2) bis (24.7) ermittelten Mittelwertquotienten und ihre entsprechenden

Streuungen zusammengestellt. Die dazu benötigten Hilfswerte lauten:

$m = 25$

Beziehung (A) in (24): $\bar{z} = \ln 10.0{,}57\,013 = 1{,}31277$

Beziehung (B) in (24): $s^2(z) = \ln^2 10.2{,}48013.10^{-2}$

$= 13{,}14\,940.10^{-2}$

Die Korrekturglieder in den geschweiften Klammern von (24.4) und dem entsprechenden Streuungsansatz $\sigma^2(q)$ sind ersichtlich unbedeutend, so dass wir sie vernachlässigen können.

Tabelle 3a:

Hypothesen und Ansätze	Q	σ(Q)	log Q	σ (log Q)
1) q *normalverteilt:* (24.2) und (24.3)	3,98065	0,33476		
2) z = log q *normalverteilt:* (24.4) und (24.5)	3,96900	0,29730		
(24.6) und (24.7)	3,96900	$\sigma_L = 1{,}0779$ $\sigma_1 = 0{,}3090$ $\sigma_2 = 0{,}2867$	*0,59868*	*0,03 256*

Das „direkt" geschätzte Q nach (24.2) und der „indirekte" FINNEY-Schätzwert von Q nach (24.4) fallen praktisch zusammen, was auf die Richtigkeit der plausibleren zweiten Hypothese hindeutet. Nach FINNEY [11] ist der direkt ermittelte Schätzwert des Durchschnittes für kleinere $s^2(z)$ nahezu ebenso wirksam wie der wirksamste Wert (24.4), so dass sich diese zwei Werte nicht sehr voneinander unterscheiden sollten, was hier der Fall ist (siehe dazu auch 125). Auch die verschiedenen Streuungswerte bleiben in kleinen Schranken; es ist wiederum verständlich, dass die „direkte" Streuung (24.3) grösser ausfallen muss als die nach (24.5) und (24.7) berechneten FINNEY-Streuungswerte, da nach den Ausführungen in 125 die erstere einen nicht sehr wirksamen Schätzwert für $\sigma^2(Q)$ darstellt.

Da wir nun die zweite Hypothese als die eindeutig bessere erkannt haben, werden für unser Portefeuille die folgenden Werte der Wirklichkeit am nächsten kommen:

$$Q = 3{,}97 \text{ mit } \quad \sigma_L(Q) = 1{,}078$$

entsprechend: $\log Q = 0{,}59879 \quad \sigma(\log Q) = 0{,}03262$

$$\sigma^2(\log Q) = 0{,}00106$$

In Analogie zu den analytischen Lösungsansätzen des 1. Kapitels wurde noch eine *3. plausible Hypothese* untersucht, nämlich die, dass $z_j^i = \log q_j^i$ *asymptotisch* PEARSON *III-verteilt* ist gemäss (13.1). Die Parameter α und γ wurden nach den in 134 erörterten Prinzipien (Vergleich der Momente, wahrscheinlichste Parameter) bestimmt, und daraus wurde der Durchschnitt Q nach (13.15) und (13.16) gewonnen. Für die Streuung $\sigma^2(Q)$ wurde eine vom Verfasser entwickelte, recht komplizierte Formel*) verwendet (siehe dazu auch die Bemerkung in 135). Die umfangreichen Rechnungen lieferten mit $\alpha = 22{,}988$ und $\gamma = 13{,}106$:

$$Q = 3{,}987 \qquad \sigma(Q) = 0{,}3334$$

Der Vergleich mit den Zahlenwerten in Tabelle 3a zeigt uns, dass auch diese 3. Hypothese aufrechterhalten werden kann. Da hierbei die numerische Auswertung sehr umständlich ist, und da wir fast dieselben Werte wie bei den anderen einfacheren Annahmen erhalten haben, wurde in 24 nicht näher darauf eingegangen. Immerhin haben wir damit einen weiteren Hinweis erhalten, dass die den Stichprobengrössen $z_i^i = \log q_j^i$ zugrunde liegende Häufigkeitsverteilung sehr gut durch die Normalverteilung approximiert wird, da der obige, für Schiefe und Exzess nach (13.8) und (13.9) verantwortliche γ-Wert relativ gross ist; demnach unterscheidet sich diese PEARSON III-Verteilung für die Grundgesamtheit nicht wesentlich von einer Normalverteilung.

32 DIE NUMERISCHE BESTIMMUNG DES DURCHSCHNITTLICHEN SCHADENBETRAGES M^1 UND DESSEN STREUUNG $\sigma^2(M^1)$

Als *Beispiel* wollen wir M^{1949} und M^{1950} ermitteln. Die theoretischen Ansätze finden sich in 25. Die Zahlenwerte bezüglich $\log y_G^i$ und $\sigma^2(\log y_G^i)$ sind der Anhangtabelle 1 unten, diejenigen von $\log Q$ und $\sigma^2(\log Q)$ dem vorangehenden Abschnitt 31 zu entnehmen.

---

*) $\sigma^2(Q) = \frac{Q^2\, 2\gamma(1+\gamma)}{m} \left\{ \left[\ln\left(1 - \frac{1}{\alpha \log e}\right) + \frac{1}{\alpha \log e - 1} \right]^2 + \frac{1}{(\alpha \log e - 1)^2\, 2(1+\gamma)} \right\}$

1949:	*1950:*
nach 31 und Anhangtabelle 1:	
$n = 737$	$n = 968$
$E(\log y_G) = 2{,}32678$	$E(\log y_G) = 2{,}32333$
$\sigma^2(\log y_G) = \frac{0{,}32065}{737} = 4{,}351.10^{-4}$	$\sigma^2(\log y_G) = \frac{0{,}34869}{968} = 3{,}602.10^{-4}$
nach 31:	
$E(\log Q) = 0{,}59879$	
$\sigma^2(\log Q) = 10{,}641.10^{-4}$	
nach (25.1) und (25.2):	
$\log M^{1949} = 2{,}92557$	$\log M^{1950} = 2{,}92212$
$\sigma^2(\log M^{1949}) = 14{,}992.10^{-4}$	$\sigma^2(\log M^{1950}) = 14{,}243.10^{-4}$
$\sigma(\log M^{1949}) = 3{,}872.10^{-2}$	$\sigma(\log M^{1950}) = 3{,}774.10^{-2}$
nach (25.3) und (25.4):	
$M^{1949} = 842{,}5$	$M^{1950} = 835{,}8$
zum Vgl.:	
$\bar{y}^{1949} = 680{,}2$	$\bar{y}^{1949} = 856{,}8$
$\sigma_L(M^{1949}) = 1{,}09325$	$\sigma_L(M^{1950}) = 1{,}09080$
Wahrscheinlichkeitsgrenzen nach (25.5) und (25.6)	
$68\%: 770{,}6 < M^{1949} < 921{,}1$	$766{,}2 < M^{1950} < 911{,}7$
$95\%: 705{,}0 < M^{1949} < 1006{,}8$	$702{,}4 < M^{1950} < 994{,}6$

Trotz des verhältnismässig umfangreichen Beobachtungsmaterials fallen demnach die Zufallsschwankungen von M^1 noch erheblich aus, was hauptsächlich auf die relativ grosse Streuung von Q, das aus 3775 Schäden bestimmt wurde, zurückzuführen ist. Wenn wir aber berücksichtigen, dass wir bei einer direkten Streuungsberechnung von $\bar{y}^1$ aus einem *Jahresbestand* praktisch wertlose Wahrscheinlichkeitsaussagen erhielten, und dass wir mit Hilfe unserer neuen Methode

durch eine *beliebige* Vermehrung des Beobachtungsmaterials die Streuung des zeitinvarianten Mittelwertquotienten Q stark reduzieren könnten, so rücken die gewonnenen Resultate in ein besseres Licht. Man erkennt nun auch, dass zu einer zuverlässigen Schätzung der Zufallsschwankungen Erfahrungsmaterial von mehreren tausend Schäden unbedingt nötig ist.

33 BESTIMMUNG EINES SCHADENSATZES ALS WEITERE ANWENDUNG

In allen Sach-, Personen- und Vermögensversicherungszweigen wird zur Beurteilung des Geschäftsergebnisses eines Jahres innerhalb einer bestimmten Risikoklasse der sogenannte „*Schadensatz* R^i" verwendet:

$$R^i = \frac{\text{Total der Schäden}}{\text{Total der Bruttoprämien}} = \frac{S^i}{\pi^i}$$

Vor allem bei Haftpflichtversicherungen ist es nicht möglich, die gesamte Schadensumme S^i am Ende des Jahres i genau anzugeben, da ein grosser Teil der angemeldeten Schäden noch unerledigt bleibt; dies ist bei den besonders ins Gewicht fallenden grossen Schäden der Fall, für die dann die geschätzte Schadenreserve eingesetzt wird. Die Kenntnis von M^i kann auch hier zu einer genaueren Bestimmung von R^i sehr gute Dienste leisten. Es wird doch:

$$R^i = \frac{n^i M^i}{\pi^i} \qquad n^i\text{: Zahl der Schäden im Jahre i.}$$

Ammeter [21] hat gezeigt, dass die *Häufigkeitsfunktion der Schadenfälle* $f(n^i)$ einer bestimmten Risikoklasse durch die Poisson-Verteilung dargestellt werden kann; diese hängt bekanntlich nur von dem einen Parameter n^i ab, ist also durch die erwartungsmässige Anzahl der Schadenfälle vollständig bestimmt. Für grössere n^i geht diese Verteilung in die *Gauss'sche Normalverteilung* über mit:

$$E(n^i) = n^i$$
$$\sigma(n^i) = \sqrt{n^i}$$

Wir können wiederum im Hinblick auf die multiplikative Zusammensetzung von R^i aus n^i und M^i annehmen, dass $\log n^i$ *näherungsweise normalverteilt* ist mit:

$$E(\log n^i) = \log n^i$$

$$\sigma(\log n^i) = \log \sigma_L(n^i) = \log\left(\frac{n^i + \sqrt{n^i}}{n^i}\right) \sim \log\left(\frac{n^i}{n^i - \sqrt{n^i}}\right)$$

n^i selbst ist dann *asymptotisch lognormalverteilt*. Der Fehler, den wir bei dieser Annahme begehen, ist unwesentlich, da $\sigma^2(\log n^i)$ sehr klein ist (siehe unten und Bemerkung in 25). Π^i ist eine feste Grösse, also nicht zufällige Variable.

Beispiel:

$n^{1949} = 737$

$\Pi^{1949} = 1'243'874.-$

$\log \Pi^{1949} = 6{,}09478$

$\log S^{1949} = 5{,}79304$

$S^{1949} = 620'929.-$

Schadensatz R^{1949}:

$\log M = 2{,}92557$ } aus 32

$\delta^2(\log M) = 14{,}992.10^{-4}$ }

$\log n = 2{,}86747$

$$\sigma_L(n) = \underbrace{\frac{764{,}148}{737}}_{1{,}03684} = \underbrace{\frac{737}{709{,}852}}_{1{,}03824} = 1{,}038$$

$\sigma(\log n) = 0{,}01620$

$\sigma^2(\log n) = 2{,}624.10^{-4}$

$$\sigma^2(\log S) = \sigma^2(\log M) + \sigma^2(\log n) = 17{,}616.10^{-4}$$

$$\sigma(\log S) = 4{,}19714.10^{-2} = \sigma(\log R)$$

$$\log R = 0{,}69286 - 1$$

$$R = 49{,}919\%$$

$$\sigma_L(R) = 1{,}1017$$

Wahrscheinlichkeitsgrenzen:

68%: $45.311\% < R^{1949} < 54{,}996\%$

95%: $41{,}130\% < R^{1949} < 60{,}587\%$

34 ZUSAMMENFASSUNG UND AUSBLICK

Allen unseren Untersuchungen lag ein bestimmtes Autohaftpflicht-Portefeuille zugrunde. Die entwickelte *Theorie* lässt sich aber ohne weiteres auf andere *Risikoklassen* (Motorräder, Lastwagen) *und*

sogar auf andere Versicherungszweige ausdehnen. Das geometrische Mittel, als charakteristische Normierungsgrösse, dürfte hauptsächlich bei allen Versicherungen auf erstes Risiko (alle Haftpflichtversicherungsarten, Heilungskosten in der Unfallversicherung) dem Schadendurchschnitt überlegen sein, also immer dort, wo wir eine grosse Variationsbreite der Schadenbeträge antreffen. Selbst bei Summenversicherungen wie z. B. in der Feuerversicherung dürfte es aber auch in Frage kommen und einer näheren Untersuchung wert sein.

Wesentliche Voraussetzung für die Anwendbarkeit unserer Theorie auf andere Gebiete (Risikoklassen bezw. Versicherungszweige) ist, dass der von uns definierte *Mittelwertquotient innerhalb einer bestimmten Risikoklasse im Laufe der Zeit konstant* bleibt. Damit kann auch bei kleinen Schadenzahlen pro Geschäftsperiode das Beobachtungsmaterial durch Zusammenfassung mehrerer Perioden vergrössert und der Mittelwertquotient zuverlässiger ermittelt werden. Dadurch können für die Zufallsschwankungen der für die einzelnen Jahre charakteristischen Grössen wie durchschnittlicher Schadenbetrag oder Schadensatz nützliche Fehlerschranken gefunden werden. Wir gewinnen damit gleichzeitig auch die Grundlagen zur Bestimmung einer angemessenen Prämie, die sich bekanntlich aus der Schadenhäufigkeit und dem durchschnittlichen Schadenbetrag zusammensetzt.

Die *analytischen Betrachtungen* mit Hilfe der Lognormalverteilung ermöglichen uns nützliche Aussagen über die Streuungen von multiplikativ zusammengesetzten Zufallsgrössen. Darin ist sie der das additive Prinzip bevorzugenden Normalverteilung überlegen.

Wir haben bei unseren Untersuchungen bewusst die Fragen der Bestimmung der Schadenreserven sowie risikotheoretische und rückversicherungstechnische Fragen nicht berührt. So dürfte es in der kollektiven Risikotheorie für praktische Anwendungen zweckmässig sein, an Stelle des Durchschnittes das geometrische Mittel als Einheit für die Schadensummen zu verwenden, da dieses, wie unsere Untersuchungen gezeigt haben, viel zuverlässiger ermittelt werden kann. Unsere *expliziten Ansätze* für die Verteilung der Schadensummen könnten vielleicht gerade für die Lösung solcher Probleme wertvoll werden.

Tabelle 1: Verteilung der Autohaftpflicht-Schäden 1950 und 1949

f_j: beobachtete Häufigkeiten
φ_j: berechnete Häufigkeiten

Klassengrenzen $x_k - x_{k+1}$ (1)	Klassenmitte x_j (2)	Entsprechende Klassengrenzen Fr. $y_k - y_{k+1}$ (3)	Klassenmitte Fr. y_j (4)	f_j 1950 (5)	φ_j 1950 nach (121.1) (6)	φ_j 1950 nach (131.1) (7)	f_j 1949 (8)	φ_j 1949 nach (131.1) (9)
0,417–0,583	0,5	2,61– 3,82	3,16	0				
0,583–0,75	0,667	3,83– 5,61	4,64	1 } 2	4,4	1,02		0,44
0,75 –0,917	0,833	5,62– 8,24	6,81	1				
0,917–1,083	1,0	8,25–12,11	10,00	6	8,7	4,04	4	2,00
1,083–1,25	1,167	12,12– 17,77	14,68	8	15,9	11,53	6	7,49
1,25 –1,417	1,333	17,78– 26,09	21,54	26	26,6	25,46	13	16,70
1,417–1,583	1,5	26,10– 38,30	31,62	34	41,2	46,13	26	32,19
1,583–1,75	1,667	38,31– 56,22	46,41	83	58,9	72,07	50	51,49
1,75 –1,917	1,833	56,23– 82,53	68,13	75	77,7	91,01	70	69,78
1,917–2,083	2,0	82,54–121,1	100,00	101	94,3	105,32	92	83,48
2,083–2,25	2,167	121,2– 177,7	146,8	120	109,6	110,65	97	88,75
2,25 –2,417	2,333	177,8– 260,9	215,4	129	113,3	110,55	97	86,38
2,417–2,583	2,5	261,0– 383,0	316,2	106	102,2	99,14	71	76,85
2,583–2,75	2,667	383,1– 562,2	464,1	84	91,9	76,22	62	64,12
2,75 –2,917	2,833	562,3– 825,3	681,3	62	74,6	63,76	54	49,87
2,917–3,083	3,0	825,4– 1211	1000,0	46	55,9	48,33	28	36,97
3,083–3,25	3,167	1212– 1777	1468	25	38,7	36,02	18	25,95
3,25 –3,417	3,333	1778– 2609	2154	20	24,6	23,98	23	17,37
3,417–3,583	3,5	2610– 3830	3162	8	14,4	16,04	2	11,25
3,583–3,75	3,667	3831– 5622	4641	13	7,8	10,45	5	6,97
3,75 –3,917	3,833	5623– 8253	6813	7		6,78	8	4,21
3,917–4,083	4,0	8254–12115	10000	1			3	2,45
4,083–4,25	4,167	12116– 17782	14678	3			5	1,39
4,25 –4,417	4,333	17783– 26100	21544	5 } 20	7,45		3	0,77
4,417–4,583	4,5	26101– 38311	31623	1		9,51	—	
4,583–4,75	4,667	38312– 56233	46415	2			—	0,86
4,75 –4,917	4,833	56234– 82539	68129	0			—	
4,917–5,083	5,0	82540–121150	100000	1			—	
				n = 968	968,15	968,01	n = 737	737,23

Hilfswerte zu Tabelle 1:

$$\bar{x}' = \frac{1}{n} \sum_{1}^{l} f_j x_j$$

$$s'^2 = \frac{1}{n-1} \sum_{1}^{l} f_j (x_j - \bar{x}')^2$$

$$\bar{y} = \frac{1}{n} \sum_{1}^{l} f_j y_j$$

(für Schäden über Fr. 8253.— wurden genaue Werte berücksichtigt.)

	1949	1950
$\bar{x}'$	2,32618	2,32333
s'^2	0,32065	0,34869
s'	0,56626	0,59050
$\bar{y}$	680,2	856,8

$$\chi^2 = \sum_{1}^{p} \frac{(f_j - \varphi_j)^2}{\varphi_j};$$

p = 19
N = Zahl der Freiheitsgrade = p − 2 = 17

Kol.(5) und (6): $\chi^2 = 58{,}7$ $P(\chi^2) < 0{,}001$
Kol.(5) und (7): $\chi^2 = 25{,}9$ $P(\chi^2) = 0{,}081$

Tabelle 2 Verteilung der Autohaftpflicht-Schäden 1938, 1947–1951. Homogenitätsteste

Klassengrenzen $Y_k = \frac{y_k}{y_G}$	*Absolute beobachtete Häufigkeiten* f_k^i						*Total* 1938; 1947–1951
	1938	1947	1948	1949	1950	1951	
0,018– 0,027	—	—	—	—	1	2	3
0,028– 0,039	—	—	5	—	1	2	8
0,040– 0,058	1	4	3	4	6	2	20
	6	8	15	10	16	12	67
0,059– 0,084	5	4	7	6	8	6	36
0,085– 0,124	6	14	14	13	26	18	91
0,125– 0,182	21	22	30	26	34	32	165
0,183– 0,268	35	42	64	50	83	63	337
0,269– 0,393	43	66	73	70	75	100	427
0,394– 0,578	43	60	74	92	101	111	481
0,579– 0,845	47	66	82	97	120	107	519
0,846– 1,234	43	57	87	97	129	123	536
1,235– 1,812	44	49	65	71	106	98	433
1,813– 2,680	31	60	48	62	84	88	373
2,681– 3,934	21	30	56	54	62	62	285
3,935– 5,783	19	29	27	28	46	45	194
5,784– 8,446	15	17	37	18	25	39	151
8,447– 12,40	11	18	14	23	20	20	106
12,41 – 18,20	4	5	11	2	8	14	44
18,21 – 26,80	5	8	11	5	13	20	62
26,81 – 39,34	1	2	3	8	7	2	23
39,35 – 57,83	3	3	1	3	1	4	15
57,84 – 84,46	1	2	3	5	3	4	18
84,87 –124,0	1 (10)	— (8)	4 (15)	3 (19)	5 (20)	3 (17)	16 (89)
124,1 –182,0	3	1	—	—	1	—	5
182,1 –268,0	1	—	2	—	2	2	7
268,1 –393,0	—	—	2	—	—	2	4
393,1 –578,0	—	—	—	—	1	—	1
	404	559	723	737	968	969	4360

Tabelle 3. Stichprobendurchschnitte und Stichproben-Mittelwertquotienten.

25 Stichproben aus dem gesamten Beobachtungsmaterial 1938, 1947–1951.

Jahr i	Stichproben j = 1, 2, .., m m=25			geometr. Mittel des Jahres i		Stichproben-Mittelwert-quotienten	
	Umfang m_j	Durchschnitt $\bar{y}_j^i$	$\log \bar{y}_j^i$	y_G^i	$\log y_G^i$	q_j^i	$\log q_j^i$
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1938	151	387,11	2,58783	144,4	2,15965	2,6803	0,42818
	145	850,11	2,92948			5,8861	0,76983
	149	578,29	2,76214	207,3	2,31664	2,7390	0,43760
1947	143	707,58	2,84978			3,3515	0,52524
	147	593,73	2,77359			2,8122	0,44905
	153	1249,27	3,09666	214,3	2,33103	5,9172	0,77212
1948	160	1076,73	3,03211			5,1000	0,70757
	159	598,08	2,77676			2,8328	0,45222
	169	1203,81	3,08056			5,7018	0,75602
	151	556,06	2,74512	212,2	2,32678	2,6338	0,42058
1949	157	796,71	2,90130			3,7736	0,57676
	137	728,61	2,86250	211,1	2,32454	3,4511	0,53796
	153	597,22	2,77614			2,8287	0,45160
	152	893,50	2,95109	210,5	2,32333	4,2321	0,62655
	138	458,68	2,66151			2,1725	0,33697
1950	158	693,02	2,84074			3,2825	0,51620
	140	2038,50	3,30931			9,6554	0,98477
	150	869,36	2,93920			4,1177	0,61466
	156	520,38	2,71632			2,4648	0,39178
	158	602,40	2,77989	229,9	2,36154	2,6203	0,41835
	153	726,41	2,86119			3,1597	0,49965
1951	157	1105,08	3,04339			4,8068	0,68185
	145	991,94	2,99648			4,3146	0,63494
	151	1345,51	3,12889			5,8526	0,76735
	143	719,38	2,85696			3,1291	0,49542
Total	3775					99,5162	14,25322
Mittel	151					$\bar{q}=3,9807$	$\bar{z}'=0,57013$
Streuung						$s^2=2,8016$	$s'^2=2,48013.10^{-2}$
M. q. Abw.						$s=1,6738$	$s'=0,15748$
Grundges. 3775		860,06	2,93453	220,8	2,34397	3,8954	0,59056

5 Literaturverzeichnis

Abkürzungen

MVM	= Mitteilungen der Vereinigung schweiz. Versicherungsmathematiker
BAF	= Bulletin de l'Institut des actuaires français
J.R.S.S.	= Journal of the Royal statistical society
Atti Ist. Naz. Ass.	= Atti dell'istituto nazionale delle assicurazioni
Ann. math. stat.	= The annals of mathematical statistics
IAK	= Internationaler Aktuar-Kongress

A. Lehrbücher

[A] Kendall M. G., The advanced theory of statistics, Vol. I and II, London, 1947.

[B] Cramer H., Mathematical methods of statistics, Princeton 1946.

[C] Depoid P., Applications de la Statistique aux assurances accidents et dommages, Paris, Berger-Levrault 1950.

[D] Riebesell P., Einführung in die Sachversicherungsmathematik, Berlin 1936.

[E] Fisher R. A., Contributions to mathematical statistics, New York 1950.

B. Einzelarbeiten

[1] Riebesell P., Fragen aus der Feuer- und Sachversicherung. Berichte der 11. Internat. Welttagung für Versicherungswissenschaft. Paris 1937, Bd. II, S. 517.

[2] Jecklin H., Ist die Anwendbarkeit der Wahrscheinlichkeitstheorie in der Lebensversicherung besser fundiert als in der Sachversicherung? Berichte 12. IAK, Luzern 1940, Bd. I, S. 377.

[3] Leepin A., Über die Anwendbarkeit von Durchschnittsverfahren zur Bestimmung der Schadenreserve in der privaten Unfallversicherung. MVM Bd. 44, Heft 2, S. 307.

[4] Myslivec V., Paretos Funktion und ihre Anwendung zur rationellen Berechnung der Versicherungstarife in der Automobilversicherung. Archiv für mathematische Wirtschafts- und Sozialforschung. Band V, Heft 1, 1939, S. 51.

[5] Pellegrin M., Tarification de l'assurance automobile. BAF Bd. 47, No. 182, Mars 1948. S. 19.

[6] Riebesell P., Einführung in die Sachversicherungsmathematik, Berlin 1936, S. 45.

[7] Lethay W., Das Gesetz über die Verteilung der Auto-Haftpflichtschäden. Das Versicherungsarchiv, Wien, Jahrgang 7, Nr. 6 (Sammelband 1936/37, S. 455).

[8] Everts H., Der Schadenverlauf in der Kraftfahrzeughaftpflichtversicherung. Neumanns Zeitschrift für Versicherungswesen, Jahrgang 58, Nr. 43, 45, 46 (Sammelband 1935, S. 1105, 1161, 1189).

[9] Konau W., Die vollständige Kalkulation der Prämien in der Sachversicherung unter Berücksichtigung der Rückdeckung. Archiv für mathem. Wirtschafts- und Sozialforschung. Band III, Heft 3, 1937, S. 192.

[10] Pae-Tsi-Yuan, On the logarithmic frequency distribution and the semi-logarithmic correlation surface. Ann. math. stat. Band 4, 1933, S. 30.

[11] Finney D. J., On the distribution of a variate whose logarithm is normally distributed. Suppl. J. R. S. S. Vol. VII, 1940–41, No. 1–2, S. 155.

[12] Cohen A. C. jr., Estimating parameters of logarithmic-normal distributions by maximum-likelihood. Journal of the American Statistical Association, Vol. 46, 1951, S. 206.

[13] Gebelein H., Logarithmische Normalverteilungen und ihre Anwendungen. Mitteilungsblatt für mathem. Statistik, Jahrgang 2, Heft 3, 1950, S. 155.

[14] d'Addario R., Sulla rappresentazioni analytica delle curve di frequenza. Atti Ist. Naz. Ass. Vol. VIII, 1936, S. 95.

[15] Amoroso L., Curve di frequenza nelle assicurazioni d'infortuni e di responsibilità civile. Le Assicurazioni d'Italia, 1932, ebenso: Atti del congresso internationale dei matematici, Vol. II, Zurigo, 1932.
La rappresentazione analitica delle curve di frequenza nei sinistri di infortuni e di responsibilità civile. Le Assicurazioni d'Italia, 1934, ebenso: Atti X IAK Rom 1934, Bd. IX, S. 44.

[16] d'Addario R., Curve di frequenza nelle assicurazioni di infortuni e di responsabilità civile. Atti del secondo congresso nazionale di scienza delle assicurazioni, Vol. III, ebenso: Rivista italiana di statistica, economia e finanza, Bd. V, No. 1, 1933.
Weiteres: Il calcolo della riserva-sinistri nella assicurazioni elementari. Giornale dell'Istituto italiano degli attuari, Bd. IV, Nr. 3, 1933.
Considerazioni intorno alla riserva-sinistri, „Assicurazioni" Bd. I, Nr. 1, 1934.
Osservazioni sulla tariffa nelle assicurazioni-danni Atti Ist. Naz. Ass. Vol. IX, 1937, S. 193.
ebenso: Berichte IAK Paris 1937, Bd. II.

[17] Risser M. R., A propos de quelques applications relative d'une part à la recherche de la loi de fréquence de certaines séries statistiques et d'autre part à l'emploi d'une transformation simple à l'étude de l'assurance accidents. BAF Bd. 42, No. 165, juin 1936, S. 245.

[18] Henry M., Le coût moyen des sinistres en responsabilité civile automobile. BAF Bd. 43, No. 169, juin 1937, S. 113.

[19] Elderton W. P., Frequency curves and correlation, London, 1927, S. 47 and 90.

[20] Pearson K., Tables of the Incomplete Γ-function, edited by Karl Pearson, Biometrika Office, University College, London.

[21] Ammeter H., Die Elemente der kollektiven Risikotheorie von festen und zufallsartig schwankenden Grundwahrscheinlichkeiten. MVM Bd. 49, Heft 1, S. 35.

LEBENSLAUF

Ich wurde am 16. September 1927 in Bern geboren. Nach Besuch der Primarschule und des Progymnasiums trat ich in die Literaturabteilung des städtischen Gymnasiums Bern über, wo ich im Herbst 1946 die Maturitätsprüfung bestand.

Seit dem Wintersemester 1946/47 bin ich mit Ausnahme eines Studienjahres in Paris an der philosophisch-naturwissenschaftlichen Fakultät der Universität Bern immatrikuliert, wo ich mich insbesondere dem Studium der Versicherungslehre, der Mathematik und der Physik widmete.

Gedruckt in Westdeutschland